

日 本 国 特 許  
JAPAN PATENT OFFICE

CERTIFIED COPY OF  
PRIORITY DOCUMENT

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2001年 4月25日

出 願 番 号

Application Number:

特願2001-127639

出 願 人

Applicant(s):

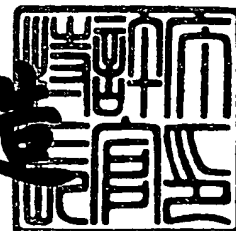
株式会社日立製作所  
日立コンピュータ機器株式会社



2001年11月30日

特許庁長官  
Commissioner,  
Japan Patent Office

及川耕造



500.41075X00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s): Y. SHIMA  
Serial No.: Not assigned  
Filed: January 10, 2002  
Title: FORM IDENTIFICATION METHOD  
Group: Not assigned



LETTER CLAIMING RIGHT OF PRIORITY

Honorable Commissioner of  
Patents and Trademarks  
Washington, D.C. 20231

January 10, 2002

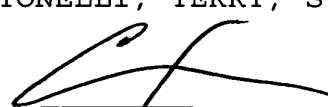
Sir:

Under the provisions of 35 USC 119 and 37 CFR 1.55, the applicant(s) hereby claim(s) the right of priority based on Japanese Application No.(s) 2001-127639 filed April 25, 2001.

A certified copy of said Japanese Application is attached.

Respectfully submitted,

ANTONELLI, TERRY, STOUT & KRAUS, LLP

  
\_\_\_\_\_  
Carl I. Brundidge  
Registration No. 29,621

CIB/amr  
Attachment  
(703) 312-6600

【書類名】 特許願  
【整理番号】 CU343  
【提出日】 平成13年 4月25日  
【あて先】 特許庁長官殿  
【国際特許分類】 G06K 9/00  
【発明者】  
    【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目280番地 株式会社  
                    日立製作所 中央研究所内  
    【氏名】 嶋 好博  
【特許出願人】  
    【識別番号】 000005108  
    【氏名又は名称】 株式会社 日立製作所  
【特許出願人】  
    【識別番号】 000233033  
    【氏名又は名称】 日立コンピュータ機器株式会社  
【代理人】  
    【識別番号】 100078134  
    【弁理士】  
    【氏名又は名称】 武 顕次郎  
    【電話番号】 03-3591-8550  
【手数料の表示】  
    【予納台帳番号】 006770  
    【納付金額】 21,000円  
【提出物件の目録】  
    【物件名】 明細書 1  
    【物件名】 図面 1  
    【物件名】 要約書 1  
【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 帳票識別方法及び帳票登録方法

【特許請求の範囲】

【請求項1】 予め登録された複数の帳票の特徴と入力された帳票の特徴とを、位置座標を基に照合して帳票の種類を識別する帳票識別方法であって、入力帳票から、帳票内に含まれる1または複数の枠、罫線または文字行を抽出し、帳票の特徴として枠、罫線または文字行の中心点座標を入力特徴として抽出し、前記入力特徴あるいは予め登録されている帳票の登録特徴のいずれかを90度単位で回転させ、枠、罫線または文字行から検出した微小傾き角に基づいて、前記入力特徴を微小角度補正し、入力特徴と登録特徴とを照合し、最も一致した角度を入力帳票の向きとし、最も一致した登録特徴を持つ帳票種類を帳票識別結果とすることを特徴とする帳票識別方法。

【請求項2】 前記最も一致した角度を持つ入力帳票の向きを出力することを特徴とする請求項1記載の帳票識別方法。

【請求項3】 登録された帳票種毎あるいは帳票の枠毎に枠の線種を有効・無効にする識別制御情報を備え、入力帳票から枠線の種類を抽出し、予め登録されている帳票の枠線の種類との一致を求め、枠の形が同一であっても枠線の種類が異なる帳票を別の種類の帳票として識別する制御、及び、枠の形が同じであれば枠線の種類が異なっても同じ種類の帳票として識別する制御を切り替えることを特徴とする帳票識別方法。

【請求項4】 予め登録された複数の帳票の特徴と入力された帳票の特徴とを、位置座標を基に照合して帳票の種類を識別する帳票識別方法であって、入力帳票の入力特徴として抽出した各位置を、予め登録されている帳票の登録特徴の基準位置から、横方向及び縦方向に一定幅で平行移動させながら前記基準位置と照合し、最も登録特徴と一致した平行移動位置を入力帳票の位置とし、最も一致した登録特徴を持つ帳票種類を帳票識別結果とすることを特徴とする帳票識別方法。

【請求項5】 前記最も登録特徴と一致した平行移動位置を入力帳票の位置として出力することを特徴とする請求項4記載の帳票識別方法。

【請求項6】 予め登録された複数の帳票の特徴と入力された帳票の特徴とを、位置座標を基に照合して帳票の種類を識別する帳票識別方法であって、入力帳票の入力特徴として抽出した各位置座標を、一定倍率ずつ横方向及び縦方向に、あるいは、横方向または縦方向に複数個の伸縮率を仮定し、前記仮定した伸縮率によって入力特徴を伸縮させ、予め登録されている帳票の登録特徴に最も一致した伸縮率を入力帳票の伸縮率とし、最も一致した登録特徴を持つ帳票種類を帳票識別結果とすることを特徴とする帳票識別方法。

【請求項7】 前記最も一致した伸縮率を入力帳票の伸縮率として出力することを特徴とする請求項6記載の帳票識別方法。

【請求項8】 予め登録された複数の帳票の特徴と入力された帳票の特徴とを、位置座標を基に照合して帳票の種類を識別する帳票識別方法であって、入力帳票の特徴に点座標を付与した入力点と、登録帳票の特徴に点座標を付与した辞書点とを備え、前記入力点と辞書点との距離を算出し、前記距離が所定値以下であれば入力点に対応付けられた辞書点が存在すると判定し、対応付けられた点数、入力点数及び辞書点数に基づいて、入力帳票の特徴と登録帳票の特徴とを照合して照合の一致の程度を求めて、入力帳票の種類を識別することを特徴とする帳票識別方法。

【請求項9】 前記入力点、辞書点は、帳票内に含まれる1または複数の枠、罫線または文字行の中心点であることを特徴とする請求項8記載の帳票識別方法。

【請求項10】 予め登録された複数の帳票の特徴と入力された帳票の特徴とを照合して帳票の種類を識別する帳票識別のために、帳票の特徴を予め登録する帳票登録方法であって、登録対象となった入力帳票画像と帳票識別により類似すると判定された登録済みの帳票画像とを画面に表示し、前記入力帳票が既に登録済みか、未登録であるかの使用者の判断情報を入力させ、新規帳票として対象帳票を登録するため対象帳票のデータを編集、印刷して再度登録試行を行うことを特徴とする帳票登録方法。

【発明の詳細な説明】

【0001】

## 【発明の属する技術分野】

本発明は、複数種類の帳票を読み取る場合に、帳票処理に先立って読み取る帳票の種類を識別する帳票識別方法及び識別した帳票を登録する帳票登録方法に係り、特に、帳票の向きや伸縮、傾きに対して安定して帳票の種類を識別ことを可能にした帳票識別方法及び帳票登録方法に関する。

## 【0002】

## 【従来の技術】

帳票識別方法に関する従来技術として、帳票の種類を識別するための特徴量、例えば、帳票中の文字コード、文字行、罫線、枠等を自動的に抽出し、これらの特徴量を手がかりに予め登録している帳票の特徴量と照合して帳票の種類を識別するという方法が知られている。

## 【0003】

前述した帳票識別の特徴量として罫線を利用する従来技術として、例えば、特開昭61-59568号公報等に記載された技術が知られている。この従来技術は、帳票内の横線及び縦線を用いて帳票の構造解析を行うことにより、帳票の種類を識別するというものである。

## 【0004】

また、帳票識別の特徴を点座標で表現して照合を行う従来技術として、例えば、特開昭62-184585号公報等に記載された技術が知られている。この従来技術は、点集合からなるパターンを照合する方法であって、2つの点集合の類似性を検出するため照合の程度を求め、また、横方向及び縦方向の小領域に分割されたブロック単位に照合の程度を高速に求めることにより、帳票の種類を識別するというものである。また、例えば、特開平8-255236号公報には、枠の中心の点を特徴量として、横方向及び縦方向を基底としたハッシュテーブルによって照合する方法が、さらに、特開2000-339406号公報には、帳票内の最小矩形の位置関係を行方向及び列方向の接続関係から求めて、最小矩形の属性を決定する方法が開示されている。

## 【0005】

また、特徴量として文字行の位置を利用する従来技術として、例えば、特開平

5-262100号公報等に記載された技術が知られている。この従来技術は、診療報酬請求明細書の様式を識別するため、抽出した文字行の位置に基づいて帳票識別を行うものである。

## 【0006】

さらに、帳票の伸縮や罫線のかすれ等のため、帳票内の個々の矩形を抽出することが不安定な帳票に対する帳票の識別に関する従来技術として、例えば、特開2000-306030号公報に記載された技術が知られている。この従来技術は、一致した矩形の座標を次に照合する矩形の基準点として、基準点を順次移動させて照合するというものである。

## 【0007】

従来、類似した性質をもつ画像を検索する方法として、例えば、「日本工業出版、画像ラボ、第11巻9号5頁から9頁、西暦2000年9月、著者名：武者義則、広池敦」に開示されている技術が知られている。この従来技術は、画像特徴量ベクトルを画像から抽出し、そのベクトル同士の距離においてキー画像から近い画像を収集するというものであり、画像特徴量として、赤緑青の三原色空間における色特徴量と輝度画像の明暗の移り変わる微分方向特徴量とを用いるというものである。しかし、この従来技術は、帳票画像に特有な罫線や文字行、枠の特徴に関して考慮されていないものである。

## 【0008】

また、従来、個人識別のための指紋照合方法として、例えば、特開2000-293688号公報等に記載された技術が知られている。この従来技術は、入力された指紋画像の特徴情報と予め格納された指紋画像の特徴情報とを照合すること、及び、特徴情報の何れか一方を倒立状態または直交状態に回転させ指紋を照合するというものである。しかし、この従来技術は、帳票画像に特有な罫線や文字行、枠の特徴に関しての配慮がないものである。

## 【0009】

また、従来、文書の回転角を検出する方法として、例えば、特開平6-103411号公報等に記載された技術が知られている。この従来技術は、文書を0度、90度、180度、270度回転させ、それぞれ文字認識を行い、その中で最

も正しいものの回転角を文書の向きとして決定するというものである。しかし、この従来技術は、帳票の向きを検出するために、それぞれの角度で文字認識を行っており、文字認識に処理時間がかかるという問題を有しており、しかも、帳票の種類を識別することについては考慮されていないものである。

## 【0010】

また、従来、帳票の識別と同時に帳票の90度単位の回転方向も識別する方法として、例えば、特開平7-249099号公報に記載された記述が知られている。この従来技術は、入力帳票を90度、180度、270度回転したものについて、罫線すなわち垂直線・水平線の分布を求めておき、これらを予め求めておいた垂直線・水平線の分布と照合することにより、帳票の識別と同時に入力された帳票の90度単位の回転方向も識別するというものである。しかし、この従来技術は、特徴として使用する罫線のはかすれや太りに対して不安定であり、接触する文字の並びによっては、文字ストロークを結ぶような偽の罫線が発生し、誤った帳票識別及び回転方向の識別を行ってしまうという問題があった。

## 【0011】

従来、枠線の線種を用いて帳票識別を行う方法として、例えば、特開平11-66228号公報等が開示された技術が知られている。この従来技術は、枠線の線種を判定して帳票読み取りのためのフォーマット情報を生成するというものである。また、この種の技術として、特開平11-85900号公報が開示されている従来技術は、実線と点線とを区別して帳票を識別するというものであり、さらに、識別精度の面で実線と点線とを区別することなく扱うことにより帳票を識別することを可能としたものである。しかし、この従来技術は、帳票の種類毎、帳票の枠毎の罫線の種類を利用するか否かを有効あるいは無効にするように切り替えることについて開示していないものである。

## 【0012】

従来、帳票に伸縮があっても帳票識別を可能にする方法として、例えば、特開2000-30600号公報等に記載された技術が知られている。この従来技術は、基準点を順次移動させて隣接する帳票の枠を照合することによって、帳票全体の伸縮による位置ずれによる照合誤りを防止するというものである。しかし、



この従来技術は、枠が欠落していて枠を抽出することができない場合や、偽の矩形が発生している場合に、枠毎の照合に誤りが発生するという問題点があり、また、基準位置がずれることについて考慮されていないものである。また、同種の技術として、特開 2 0 0 0 - 1 2 3 1 7 4 号公報、特開平 8 - 3 1 5 0 6 8 号公報、特開平 7 - 2 4 9 0 9 9 号公報等に記載された技術が知られている。これらの従来技術は、一定領域内に含まれる罫線の間隔を照合し、その照合結果から帳票画像の伸縮率を推定する方法に関するものであるが、罫線同士の照合において、罫線の欠落や偽の罫線の発生がある場合、照合自体が誤り、そのため、推定した伸縮率の値が誤るという問題について解決されておらず、罫線の照合に依存して伸縮に対処する方法であるため、罫線の照合が誤ると伸縮に誤りが発生するという問題がある。

## 【 0 0 1 3 】

## 【発明が解決しようとする課題】

前述した従来技術のうち、帳票識別の対象となる入力帳票の向きを識別する技術は、予め登録する帳票の隅に特殊なマークや指定の文字列を印刷しておき、入力帳票に対してそのマークなどを読み取ることにより、帳票が 9 0 度回転した向きにスキャナに設定された場合や、上下逆すなわち 1 8 0 度回転した状態の帳票がスキャナで入力された場合に、帳票の種類を識別することを可能としたもので、帳票に特殊な指定された印刷領域が必要であり、帳票のスペースを有効に使用することができないという問題点を有している。

## 【 0 0 1 4 】

また、前述した従来技術のうち、帳票の識別と同時に帳票の 9 0 度単位の回転方向も識別する方法として、前述した特開平 7 - 2 4 9 0 9 9 号公報に記載された技術は、入力帳票を 9 0 度、1 8 0 度、2 7 0 度回転したものについて、罫線すなわち垂直線・水平線の分布を求めておき、これらを予め求めておいた垂直線・水平線の分布と照合を行うことで帳票の識別と同時に入力された帳票の 9 0 度単位の回転方向も識別するという技術は、特徴として利用する罫線のかすれや太りがある場合に、識別が不安定となり、また、接触する文字の並びによっては、文字ストロークを結ぶような偽の罫線が発生し、誤った帳票識別及び回転方向の

識別を行ってしまうという問題点を有している。

【0015】

また、前述した従来技術のうち、帳票識別の対象となる入力帳票中の枠線の種類、実線と点線とを区別して帳票を識別する技術は、帳票の種類毎、帳票の枠毎の罫線の種類を利用するか否かを有効あるいは無効にするように切り替えることについて考慮されていない。このため、この従来技術は、点線の間隔が狭いような帳票の場合、点線の点が太って隣の点とつながり実線となった場合、異なった帳票と誤識別されることや、該当する帳票が検出できないとして拒絶となる恐れがあるという問題点を有し、また、帳票によっては、枠の形が同じで、実線と点線との区別が不要で、特定部分が実線であっても点線であっても、同一帳票とみなすべき帳票であるとするような帳票が混在して入力された場合、点線と実線とを全種類の帳票に対して同一とみなして帳票識別を行うか、点線と実線との区別をして異なる帳票とみなして帳票識別を行うかの、何れかの方法を択一的に選択する以外に対処することができないものである。すなわち、この従来技術は、実線と点線との区別が不要で、特定部分が実線であっても点線であっても、同一帳票とみなすべき帳票に対しても、帳票識別用の辞書や帳票読み取りのためのフォーマット情報を予め別々に登録しておく必要があり、辞書登録の作業が多くかかり、また、辞書を記憶する容量が多くなるという問題点を有している。

【0016】

さらに、前述した従来技術のうち、帳票の伸縮に対応するため基準点を順次移動させて照合するという方法は、罫線の欠落や偽罫線の出現により、照合途中で新たに設定した基準点を誤ると、帳票識別に誤りが生じ、あるいは、拒絶となる恐れがあるという問題点を有している。

【0017】

さらに、前述で説明した従来技術は、いずれも、登録する帳票の種類が多くなると、新たに登録対象となった帳票が既に登録済みか、未登録であるかを使用者が判断することが困難となり、また、登録済みの類似帳票が同一種類の帳票として識別された場合、その帳票を類似帳票と区別することができないという問題点を有している。

## 【 0 0 1 8 】

本発明の第1の目的は、前述した従来技術の問題点を解決し、帳票の向きが任意に設定されて入力された帳票、例えば、90度回転した状態や、上下逆すなわち180度回転した状態の帳票がスキャナで入力された場合であっても、正立した状態の帳票の場合と同様に、高い信頼性を持って帳票の種類を識別すると同時に帳票の回転方向を検出することができる帳票識別方法を提供することにある。

## 【 0 0 1 9 】

本発明の第2の目的は、予め登録した帳票毎あるいは当該帳票の枠毎に制御情報を備え、枠線の種類、例えば、実線あるいは点線であるということを利用して、枠の形が同じであっても枠線の種類が異なる帳票を別の種類の帳票として識別するか、あるいは、枠線の種類が異なる帳票を同じ種類の帳票として識別するかを切り替えて帳票の種類を識別することができる帳票識別方法を提供することにある。

## 【 0 0 2 0 】

本発明の第3の目的は、帳票に位置ずれがあっても、高精度に帳票の種類を識別することができる帳票識別方法を提供することにある。

## 【 0 0 2 1 】

本発明の第4の目的は、罫線のかすれ等による欠落や偽の罫線の出現に対しても、また、帳票に伸縮があっても、高精度に安定して帳票の種類を識別することができる帳票識別方法を提供することにある。

## 【 0 0 2 2 】

本発明の第5の目的は、新たに登録対象となった帳票が既に登録済みか、未登録であるかを使用者に目視で判断するための画面インタフェースと、対象帳票を編集して印刷し再度登録試行する機能と、印刷した帳票に対応した帳票識別用辞書が保管されていることを保証する機能とを備えた帳票登録方法を提供することにある。

## 【 0 0 2 3 】

## 【課題を解決するための手段】

本発明による帳票識別方法は、入力帳票から特徴として1または複数の枠の中

心点を抽出し、入力帳票の入力特徴あるいは予め登録されている登録帳票の特徴のいずれかを90度単位で回転させ、入力特徴と登録特徴と照合し、最も一致した角度を帳票の向きとして、特徴に対応する帳票種類を帳票識別結果として出力することを特徴とし、これにより、帳票が90度回転した向きにスキヤナに設定された場合や、上下逆すなわち180度回転した状態の帳票がスキヤナに入力された場合であっても、正立の状態の帳票と同様に帳票識別を行うことが可能である。特徴としての枠の中心点は、他の特徴、例えば、罫線の投影分布等と比べて、安定して抽出することができ、識別誤りを少なくすることができる。何故なら、罫線は、文字行の内部にも偽罫線として抽出されるが、4辺が罫線で構成される枠を抽出することにより、その枠の罫線を枠線として利用しているため、誤抽出をすくなくすることができるためである。

## 【0024】

また、本発明による帳票識別方法は、入力帳票から枠線の種類を抽出し、予め登録された帳票の枠線の種類との一致を求め、枠の形が同じであっても枠線の種類が異なる帳票を別の種類の帳票として識別することを可能にし、さらに、枠線の種類が異なる帳票であっても枠の形が同じであれば同じ種類の帳票であると識別することも可能にすることを特徴とする。

## 【0025】

また、本発明による帳票識別方法は、入力帳票の特徴を有する各位置を予め登録された特徴の基準位置から、横方向ならびに縦方向に一定幅で平行移動させながら照合し、最も登録特徴と一致した平行移動位置を入力帳票の位置とし、特徴に対応する帳票種類を帳票識別結果として出力することにより、帳票に位置ずれがあっても帳票識別を可能にすることを特徴とする。

## 【0026】

また、本発明による帳票識別方法は、帳票の特徴の照合と特徴の伸縮とを独立させ、特徴の照合の誤りに関係なく、特徴を伸縮させることを特徴とする。ここでは、入力帳票の特徴を有する各位置を一定倍率ずつ横方向ならびに縦方向に伸縮させ、最も登録特徴と一致した伸縮倍率を入力帳票の伸縮率として出力し、特徴に対応する帳票種類を帳票識別結果として出力することにより、帳票に伸縮が

あっても、帳票識別を可能にする。

## 【 0 0 2 7 】

本発明による帳票識別方法は、新たに登録対象となった帳票画像と帳票識別により類似すると判定された登録済みの帳票画像を画面に表示し、当該帳票が既に登録済みか、未登録であるかの使用者の判断情報を入力させ、新規帳票として対象帳票を登録するため対象帳票のデータを編集して印刷し再度登録試行することにより帳票登録を可能にすることを特徴とする。

## 【 0 0 2 8 】

本発明による帳票識別方法において、入力帳票の特徴と登録帳票の特徴を照合する方法は、入力帳票の特徴に点座標を付与した入力点と、登録帳票の特徴に点座標を付与した辞書点とを備え、前記入力点と辞書点との距離を算出し、前記距離が所定値以下であれば入力点に対応付けられた辞書点が存在すると判定し、対応付けられた点数、入力点数及び辞書点数を基に照合の一致の程度を求めることにより高精度な帳票識別を可能にする点にある。

## 【 0 0 2 9 】

## 【発明の実施の形態】

以下、本発明による帳票識別方法及び帳票登録方法の実施形態を図面により詳細に説明する。

## 【 0 0 3 0 】

図1は本発明の一実施形態による帳票識別及び帳票登録を実行する帳票識別システムの構成例を示すブロック図である。図1において、100はネットワーク、101は辞書登録ステーション、102、107はスキャナ、103は登録対象帳票画像ファイル、104、108は帳票辞書ファイル、105は表示装置、106は帳票識別ステーション、109は帳票作成ステーション、110は帳票ファイル、111は帳票プリンタである。

## 【 0 0 3 1 】

図1に示す帳票識別システムは、通信ネットワーク100を介して辞書登録ステーション101、帳票識別ステーション106、帳票作成ステーション109が接続されて構成されており、これらのステーションが相互に連携して動作する

ことが可能である。辞書登録ステーション101は、帳票の種類を識別するために予め帳票画像の特徴を登録する。辞書登録ステーション101には、帳票画像を採取するスキャナ102、登録対象となった帳票画像を保管する登録対象帳票画像ファイル103、登録した帳票画像の特徴を保管する帳票辞書ファイル104が備えられている。また、辞書登録ステーション101は、表示装置105を有し、この表示装置105に、登録済み帳票画像やスキャナで採取した新たな帳票画像をその画面に表示し、対話的に登録作業を行う。

## 【0032】

帳票識別ステーション106は、スキャナ107から入力された帳票画像の種類を、登録した帳票画像の特徴を保管する帳票辞書ファイル108を用いて識別する。帳票辞書ファイル108の内容は、辞書登録ステーション101側の帳票辞書ファイル104を複製したものであり、通信ネットワークを介して、常に最新のデータに更新することができる。なお、帳票識別に用いる帳票辞書ファイルとして、通信ネットワークを介して辞書登録ステーション側の帳票辞書ファイル104をアクセスして用いる構成でもよい。

## 【0033】

帳票作成ステーション109は、帳票ファイル110と帳票プリンタ111とを備えており、帳票ファイル110に保管された帳票データに対して、帳票プリンタ111を用い帳票を紙に印刷する。帳票作成ステーション109は、帳票識別が容易になるよう帳票データを編集して、印刷し、辞書登録ステーション101において、スキャナ102で印刷した帳票の画像を採取し新規登録の試行を行うことができる。

## 【0034】

図2は辞書登録ステーション101での帳票識別のための辞書登録の処理動作を説明するフローチャートであり、以下、これについて説明する。

## 【0035】

(1) まず、帳票画像を入力し、帳票画像の特徴として枠を抽出する。この処理は、帳票画像から横線と縦線とを抽出し、その交点を基に枠を抽出する処理であり、さらに、枠内の中心点の座標を特徴として抽出する。一般に、枠は、複数個

中種されるので、複数の枠の中心点の座標が特徴として抽出される（ステップ 200、201）。

【0036】

(2) 次に、既に登録している帳票辞書と入力帳票から抽出した特徴との照合を行い、照合の程度を求める（ステップ202）。

【0037】

(3) ステップ202での照合の程度を基に、リジェクトかアクセプトかを決定する。ここでは簡単に、照合の程度が所定値より大きければアクセプト、所定値より小さければリジェクトとしてもよい（ステップ203）。

【0038】

(4) ステップ203で、リジェクトと判定された場合、入力帳票から抽出した当該特徴を帳票辞書ファイル104に辞書として追加登録し、該当帳票画像を登録対象帳票画像ファイル103に保管する（ステップ204、205）。

【0039】

(5) ステップ203で、アクセプトと判定された場合、入力された帳票が登録済あるいは類似帳票が登録されている可能性があるとして、表示装置105に警告を表示すると共に、登録帳票画像を登録対象帳票画像ファイル103から読み込んで、入力された帳票画像と登録帳票画像とを画面に表示し、当該帳票の確認及び帳票作成ステーション109での帳票の変更を行う（ステップ206、207）。

【0040】

前述で説明したように、辞書登録ステーション101は、登録対象となった帳票画像と帳票識別により類似すると判定された登録済みの帳票画像とを画面に表示し、当該帳票が既に登録済みか、未登録であるかを使用者に判断させ、その判断情報を入力させている。これにより、新規帳票として対象帳票を登録するため対象帳票のデータを編集して印刷し再度登録試行することができる。

【0041】

図3は図2により説明した辞書登録における対話操作を行うための画面表示例を説明する図である。表示装置105の表示画面300には、入力帳票画像30

1と登録済みあるいは類似帳票である登録帳票画像304とが表示される。また、表示画面300には、表示画像の説明文303、304が表示され、また、使用者に確認の必要を指示するため、警告文305が表示されると共に、対象帳票のデータを編集して印刷する帳票作成ステーション109の処理に移るためのクリックボタン306が表示される。

## 【0042】

図4は帳票識別ステーション106での入力帳票の特徴抽出処理と帳票識別処理との動作を説明するフローチャートである。図4に示すフローにおいて、ステップ400で入力帳票画像から帳票識別用の特徴を抽出する。説明している例では、特徴として帳票内にある全ての枠の中心点の座標を抽出する。但し、特徴として枠の中心点の座標に限定するものでなく、例えば、罫線の中心点の座標や文字行の中心点の座標であってもよい。次に、ステップ401で抽出した特徴に基づいて帳票の種類を識別する。

## 【0043】

図5は図4のステップ401での帳票識別処理の動作を説明するフローチャートであり、以下、これについて説明する。

## 【0044】

(1) 図4のステップ400で抽出した入力帳票画像の特徴、すなわち、枠の中心点(入力点)の座標を90度単位で回転する。ここでは、左回り90度、右回り90度、上下逆転の180度への3通りの回転を行う。これらの回転は、特徴としての枠の中心点の座標を入れ替えることにより行うことができる(ステップ500)。

## 【0045】

(2) ステップ500での3通りの回転のそれぞれ、及び、回転させない場合について、中心座標を求め、後述するステップ501の探索照合、ステップ502の類似度登録を繰り返す(ステップ505)。

## 【0046】

(3) 探索照合の処理では、基準位置を所定幅で横方向、次いで、縦方向に平行移動しながら、入力点と辞書中の点との照合を行う。そして、類似度登録の処理



では、ステップ500で設定した回転角毎に、照合により得られた類似度を求めて、その類似度とその回転角を登録する（ステップ501、502）。

## 【0047】

(4) 次に、登録した回転角毎の類似度の中で、最大の類似度を持つ回転角を帳票の方向として決定し、さらに、その最大の類似度を持つ辞書の種類を対応する帳票の種類であると決定する（ステップ503、504）。

## 【0048】

図6は図5のステップ501での探索照合処理の動作を説明するフローチャートであり、以下、これについて説明する。

## 【0049】

(1) まず、点座標の基準位置を平行移動するための移動開始位置を設定し、次に、基準位置からの探索刻み幅、すなわち、基準位置の平行移動の単位を設定する（ステップ600、601）。

## 【0050】

(2) 基準位置のX方向を設定した刻み幅でずらし、さらに、基準位置のY方向を設定した刻み幅でずらしながら、ステップ604、605、606の処理を繰り返す（ステップ602、603）。

## 【0051】

(3) 点照合の処理において、入力点と辞書点との照合を辞書が尽きるまで行う。そして、類似度登録の処理において、当該点照合の類似度と、そのときのX方向、Y方向のずらし量とを登録する（ステップ604、605）。

## 【0052】

(4) ステップ604、605の処理において、回転と、X方向、Y方向のずらし量との複数の組み合わせ毎の点照合の類似度と、そのときのX方向、Y方向のずらし量とが登録されるので、その中から最大の類似度を持つX方向、Y方向のずらし量を持つ辞書の種類を、X方向、Y方向のずらし量と共に登録する。この登録されたX方向、Y方向のずらし量は、後に、帳票内の読み取るべき枠の一を決定し、枠内の文字等の読み取りに使用される（ステップ606）。

## 【0053】

図7は図5のステップ604での点照合処理の動作を説明するフローチャートである。図7に示すフローにおいて、まず、ステップ700で、基準位置からの位置ずらし量を設定する。次に、ステップ701で、辞書が尽きるまでステップ702での類似度の抽出を繰り返す。そして、ステップ703で抽出された類似度の大きい順に帳票種類を並べ換える。

## 【0054】

図8は図7のステップ702での類似度抽出処理の動作を説明するフローチャートであり、以下、これについて説明する。この処理は、入力帳票に伸縮があるとして、伸縮率を複数通りに仮定して類似度を抽出するものであり、照合処理と特徴に対する伸縮処理を独立に行っているため、照合処理の誤りに影響を受けないという利点を有する。

## 【0055】

(1) まず、拡大縮小の倍率を複数設定し、それらが尽きるまで、ステップ801、802、803の処理を繰り返すことを設定する(ステップ800)。

## 【0056】

(2) 対応する拡大縮小率に従って入力点座標の拡大縮小を行う。この処理で、伸縮率を複数通りに仮定して入力点座標の拡大縮小を行っているため、帳票が伸縮している場合でも類似度を算出でき、また、特徴として利用する点に欠落や偽の点の発生があっても、信頼性よく類似度を算出できるという効果がある(ステップ801)。

## 【0057】

(3) 次に、入力点と照合した辞書点の個数を投票数として求めて投票数を抽出し、投票数と入力点数及び辞書点数をもとに類似の度合いを算出する(ステップ802、803)。

## 【0058】

前述の処理において、類似度の算出方法としては、例えば、投票数を入力点数と辞書点数の和で割った百分率を用いてよい。すなわち、類似度は、投票数/(入力点数+辞書点数)の百分率となる。また、類似度は、例えば、投票数に対して入力点数と辞書点数の差を引き算した個数を入力点数と辞書点数の和で割った

百分率を用いてもよい。すなわち、類似度は、 $(投票数 - 絶対値(辞書点数 - 入力点数)) / (入力点数 + 辞書点数)$ の百分率となる。この場合、投票数に対して入力点数と辞書点数との差を引き算して求めているので、入力点が多数検出された場合、投票数を減点して類似度を低下させることができるという効果がある。

## 【0059】

図9は図8のステップ802での投票数抽出処理の動作を説明するフローチャートである。この処理は、注目する入力点に対して近傍にある辞書点の内、最小距離にある辞書点を求め、その最小距離の長さから照合する辞書点が存在するかどうかを判定すると共に、入力点に対して辞書点が存在する入力点の個数を投票数として計数する処理である。

## 【0060】

図9に示すフローにおいて、まず、ステップ900で、投票数を初期化する。そして、ステップ901で、入力点が尽きるまで、ステップ902以下の処理を繰り返すことを設定する。ステップ902、903の処理で、辞書の点が尽きるまで、入力点と辞書点との最小距離を算出する。そして、ステップ904で当該距離が所定値より小さいか否かを判定し、小さければ、ステップ905で投票数の値を1個増加させる。以上の処理を繰り返すことにより、投票数を算出することができる。

## 【0061】

図10は図4のステップ400での枠抽出処理の動作を説明するフローチャートであり、以下、これについて説明する。

## 【0062】

(1) まず、帳票画像から横線と縦線とを検出する。この処理において、画像を黒線分すなわち黒ランの集合で表現し、線分長が長い黒ランを横線として抽出する。また、縦線の抽出は、画像を90度回転して、同じく線分長が長い黒ランを抽出することにより行う。なお、別の方法として、黒画素の塊の輪郭線を抽出して枠線を取り出す方法を使用してもよい(ステップ1000)。

## 【0063】

(2) 次に枠を検出する。この枠の検出は、縦線及び横線の交点から枠を取り出すことにより行うことができる。また、詳細な枠検出方法は、特開平8-125855号公報に開示された方法であってよい(ステップ1001)。

## 【0064】

(3) 次いで、枠の4隅の座標を検出し、枠の中心点座標を算出し、この中心点座標を帳票識別のための特徴とする(ステップ1002、1003)。

## 【0065】

(4) さらに、枠線の傾きや帳票縁の傾きより検出された帳票の微小な傾き角を基に、算出した中心点座標に対して傾き補正を行う。ここでの補正は、微小な傾き角に対する点座標の補正であり、90度単位の回転に関しては、すでに説明したステップ505の処理で行われる。このように、微小な傾きに対して傾き補正を行った中心点座標が帳票識別のための特徴とされる(ステップ1004)。

## 【0066】

図11は帳票識別ステーション106での入力帳票の特徴抽出処理と帳票識別処理とその他の動作例を説明するフローチャートであり、図4により説明した例とは別の処理例である。この処理は、特徴として文字行の位置座標を用いて帳票識別を行おうとするものである。

## 【0067】

図11に示すフローにおいて、ステップ1101の処理で、入力帳票画像から文字行を抽出する。詳細な文字行抽出の方法は、特開平11-232376号公報に開示されたバーコード行の抽出処理方法を使用することができる。そして、文字行の座標、例えば、文字行内部の中心点の座標を特徴とし、これにより、ステップ1102の処理で帳票識別を行う。なお、ステップ1102の帳票識別処理は、図4のステップ401の処理と同様の処理でよい。

## 【0068】

図12は図4のステップ400での枠抽出処理について説明する図であり、帳票識別の特徴としての枠の中心点について説明する図である。図12(a)に入力帳票画像1200の例を示しており、この例では、入力帳票画像1200に、枠1201、1202、1203の3つの枠が存在している。そして、図12(

b) に入力帳票画像1200内の各枠の中心点1211、1212、1213を抽出した結果を示している。それぞれの枠内部の中心点は、横方向ならびに縦方向の座標として表現される。

## 【0069】

図13は帳票辞書ファイル104あるいは108に格納されている辞書点について説明する図である。図13(a)～図13(d)に、帳票の種類が登録済みである辞書の例を1300、1301、1302、1303として示している。帳票の種類に対応してそれぞれの辞書が具備されている。辞書1300には、辞書点1310、1311、1312、1313が位置座標として具備され、また、辞書1301には、辞書点131、1315、1316が、辞書1302には、辞書点1317、1318、1319が、辞書1303には辞書点1320、1321、1322が、それぞれ位置座標として具備されている。

## 【0070】

図14は図9のステップ903での入力点と辞書点との最小距離を算出する方法を説明する図である。図14において、1400で示した横方向X、1401で示した縦方向Yの帳票平面において、入力帳票画像から抽出した入力点（黒丸で示す）を1410、1411、1412とする。また、ある辞書の辞書点（白丸で示す）を1420、1421、1422、1423とする。

## 【0071】

いま、任意の入力点1410に対して、近傍の最短距離に存在する辞書点を求めるため、当該入力点1410と辞書点1420、1421、1422、1423との距離を算出し、その最短距離を有する辞書点、本例では、辞書点1420を求める。それぞれの入力点に対して、同様な処理を行い、最短距離にある辞書点を求める。この結果、入力点1411の最短距離にある辞書点として1421が、入力点1412に対して1422がそれぞれ抽出される。なお、入力点を基準にして対応する辞書点を検出しているため、辞書点1423については、対応する入力点はない。

## 【0072】

図15は図6のステップ602、603で示した座標値をX、Y方向へずらし

た点照合について説明する図である。図15に示す入力帳票画像において、1500で示した横方向X、1501で示した縦方向Yの原点の位置を1540とする。また、入力帳票画像から抽出した入力点を1520、1521、1522とし、ある辞書の辞書点を1530、1531、1532、1533とする。そして、縦方向の破線1502～1507は、原点の位置を横方向に一定間隔でずらすときのずらし位置を、また、横方向の破線1508～1512は、原点の位置を縦方向に一定間隔でずらすときのずらし位置を示している。

## 【0073】

図15において、原点を前述の破線の交点に平行移動させる。すなわち、入力点の座標の値を所定量だけ更新する。そして、平行移動した入力点座標に対して、図6のステップ604の点照合の処理を行う。本発明の実施形態は、このように、原点を一定間隔で横方向及び縦方向に平行移動して点同士の照合を行っているため、入力帳票に位置ずれがあっても高精度に帳票識別が可能となる。

## 【0074】

図16は枠の特徴として用いる枠線の種類について説明する図である。前述で説明した図10のフローにおける線分を抽出するステップ1000において、実線だけでなく、例えば、点線を抽出し、これを枠の特徴として帳票識別に利用することができる。枠を構成する線の種類として、実線と点線とがある場合、図16に示すように、枠の種類としては、枠1600～1615として示す16通りの種類がある。なお、枠線の種類として、実線と点線とに限定する必要はなく、例えば、一点鎖線や破線などの線種を含めてもよい。また、線の太さを線種に含めてもよい。枠の線種が異なれば、枠の形や位置が同一であっても、帳票の種類は異なる。

## 【0075】

本発明の実施形態では、枠線の種類が異なる16種類の枠にそれぞれ属性を与える。属性は、枠線の種類に対応して、例えば、1から16までの番号であってもよい。例えば、枠1600は4本の枠線が実線である。また、枠1601～1604は3本の枠線が実線で、残り1本の枠線が点線である。枠1605～1610は2本の枠線が実線で残り2本の枠線が点線である。枠1611～1614

は3本の枠線が点線で残り1本の枠線が実線である。枠1615は4本の枠線が点線であり、これらは、すべて異なる種類の枠である。なお、ステップ1000で罫線のうち点線を抽出する処理には、例えば、特開平9-319824号公報に開示された点線抽出方法を使用することができる。

## 【0076】

図17は図8のステップ802での投票数抽出の処理において、点の属性を利用する場合の処理動作を説明するフローチャートである。ここでの処理は、点の属性として枠線の線種を付与している。そして、この処理は、注目する入力点に対して近傍にある辞書点の内、入力点と辞書点との属性が同じでかつ最小距離にある辞書点を求め、その最小距離の長さから照合する辞書点が存在するか否かを判定すると共に、入力点に対して辞書点が存在する入力点の個数を投票数として計数する処理である。

## 【0077】

図17において、まず、ステップ1700で投票数を初期化する。そして、ステップ1701で、入力点が尽きるまで、ステップ1702以下の処理を繰り返すことを設定する。ステップ1702では、辞書の点が尽きるまで、ステップ1703で入力点と辞書点の属性が一致する辞書点を求め、ステップ1704で入力点と当該辞書点の最小距離を算出する。そして、ステップ1705で当該距離が所定値より小さいか否かを判定し、小さい場合に、ステップ1706で投票数の値を1個増加させる。以上の処理を繰り返すことにより、投票数を算出することができる。

## 【0078】

前述した図17の処理により、枠線の種類が異なる帳票に対して、帳票の種類が異なると判定することができる。これにより、例えば、枠線の属性が異なっても同じ種類の帳票とみなせる場合、文字読み取りのための書式データを共通化することができ、書式データ作成の作業を省略できるという利点を得ることができる。また、帳票画像において破線が画像2値化の影響によりつぶれ、破線が繋がった実線となる場合がある。このような帳票に対しては、枠線の属性を無視することにより、信頼性よく帳票の種類を識別できる。

## 【0079】

図18は図5のステップ500での枠座標の90度回転の処理について説明する図である。図18において、入力帳票画像から抽出した特徴を1800として示している。ここでは、枠の中心点の位置座標を特徴として具備している。この入力特徴1800に対して、右に90度回転した特徴が特徴1801であり、また、当該入力特徴1800に対して、上下逆転した特徴が特徴1802であり、左に90度回転した特徴が特徴1803である。本発明の実施形態は、入力特徴を前述のように90度単位で回転させ照合を行っているため、スキャナに帳票が任意の向きに設定されていても帳票の識別が可能である。なお、入力点を回転させる方法とは別に、辞書点を回転させて照合を行ってもよい。

## 【0080】

図19は図8のステップ802での投票数抽出の処理において、点の属性を有効または無効の何れかに切り替えて投票数を求める場合の処理動作を説明するフローチャートであり、以下、これについて説明する。

## 【0081】

この切り替え処理は、帳票の種類毎に実行される。但し、帳票の種類毎に切り替え処理が限定されるのではなく、帳票内の枠毎に切り替え処理が実行されてもよい。そして、予め登録された帳票毎、あるいは、帳票内の枠毎に、枠線の種類を有効・無効にする識別制御情報を具備する。ここでは、一例として、識別制御情報として属性フラグを備え、ステップ1907で用いるこの属性フラグにより制御を切り替えることとする。また、この例では、点の属性として枠線の線種を付与しており、線種を考慮して識別する処理と線種の違いを無視して識別する処理とを切り替えるものとする。そして、この処理は、注目する入力点に対して近傍にある辞書点の内、入力点と辞書点との属性が同じでかつ最小距離にある辞書点を求め、その最小距離の長さから照合する辞書点が存在するか否かを判定すると共に、入力点に対して辞書点が存在する入力点の個数を投票数として計数する処理である。

## 【0082】

(1) まず、投票数を初期化し、入力点が尽きるまで、ステップ1902以下の



処理を繰り返すことを設定する（ステップ1901）。

【0083】

(2) 辞書の点が尽きるまで、各枠の属性、例えば、線種を有効にするか無効にするかを判定する。もし、属性を有効にする場合、入力点と辞書点との属性が一致する辞書点を求め、入力点と当該辞書点と間の最小距離を算出する（ステップ1902、1907、1903、1904）。

【0084】

ここでは、前述のステップ1907で、属性を無効にした場合、ステップ1903の処理以下の最小距離算出を実行しないが、必ずしもこのように限定する必要はない。例えば、ステップ1907で属性フラグが無効の場合、ステップ1903で入力点と辞書点との属性が一致しなくても、ステップ1904の処理で最小距離の算出を実行し、属性を無視する方法でもよい。

【0085】

(3) そして、ステップ1904で算出した最小距離が所定値より小さいか否かを判定し、小さければ、投票数の値を1個増加させる。以上の処理を繰り返すことにより、投票数を算出することができる（ステップ1905、1906）。

【0086】

前述した図19に示す処理により、枠線の種類が異なる帳票に対して、帳票の種類が同一と判定する、あるいは、異なる種類と判定することを切り替えて、投票数を算出することができる。

【0087】

図20は帳票識別に用いる帳票画像の特徴例について説明する図である。帳票の特徴として、図20に示すように、帳票中の枠2001における枠内部の中心点2000の座標、帳票中の罫線2003における罫線の中心点2002の座標、帳票中の文字行に外接する矩形2005の中心点2004の座標等があり、これらを帳票識別に利用する特徴として使用することができる。

【0088】

図22は帳票画像の特徴として、図20により説明した3種の特徴、すなわち、枠、罫線、文字行矩形を用いて帳票の識別を行う処理動作の例を説明するフロ

ーチャートである。図 2 2 において、ステップ 2 2 0 0 で帳票の特徴として枠を抽出し、ステップ 2 2 0 1 で枠を利用して帳票識別を行う。次いで、ステップ 2 2 0 2 で帳票の特徴として罫線を抽出し、ステップ 2 2 0 2 で罫線を利用した帳票識別を行う。さらに、ステップ 2 2 0 4 で文字行を抽出し、ステップ 2 2 0 5 で文字行を利用した帳票識別を行う。最後に、ステップ 2 2 0 6 でステップ 2 2 0 1、2 2 0 3、2 2 0 5 での各帳票識別結果の複合判定を行う。

## 【 0 0 8 9 】

図 2 3 は図 2 2 のステップ 2 2 0 6 での帳票識別結果の複合判定の処理動作を説明するフローチャートであり、以下、これについて説明する。

## 【 0 0 9 0 】

まず、ステップ 2 3 0 0 で予め登録している帳票の種類が尽きるまで、ステップ 2 3 0 1 ~ 2 3 0 5 の処理を繰り返すことを指定する。ステップ 2 3 0 1 の処理で、複合判定のための指標値を帳票種類毎に初期化して備える。次に、ステップ 2 3 0 2 の処理で、枠利用の帳票識別結果を当該指標値、ここでは、初期値に加算する。そして、ステップ 2 3 0 3 の処理で、罫線枠利用の帳票識別結果を当該指標値、ここでは、初期値 + 枠利用の識別結果の指標値に加算する。さらに、ステップ 2 3 0 4 の処理で、文字行利用の帳票識別結果を当該指標値、ここでは、初期値 + 枠利用の識別結果 + 罫線利用の識別結果の指標値に加算する。次に、ステップ 2 3 0 5 の処理で、ステップ 2 3 0 4 で得られている初期値 + 枠利用の識別結果 + 罫線利用の識別結果 + 文字行利用の識別結果の指標値を登録する。最後に、ステップ 2 3 0 6 の処理で、指標値の大小に沿って識別結果である帳票種類を並べ替える。

## 【 0 0 9 1 】

前述において、指標値として各帳票識別候補の順位を用いた場合、指標値が小さい順に帳票種類がソートされる。あるいは、指標値として各帳票識別候補の類似度を用いた場合、指標値が大きい順に帳票種類がソートされる。本発明の実施形態は、前述のように、複数の帳票の特徴を用いて帳票識別結果を出力することができるため、高精度な帳票の識別が可能である。例えば、枠と下線とが混在するような帳票を高精度に識別することができる。

## 【0092】

図24は3種の帳票の特徴を用いた帳票識別の処理動作の他の例を説明するフローチャートであり、図23入力より説明したフローによる方法とは別の方法による例である。

## 【0093】

図24において、まず、ステップ2400で枠を抽出し、ステップ2401で枠利用の帳票識別を行う。そして、ステップ2402で枠利用の帳票識別が拒絶か否かを判定し、もし拒絶であれば、ステップ2403で罫線を抽出し、ステップ2404で罫線利用による帳票識別を行う。そして、ステップ2405で罫線利用の帳票識別が拒絶か否かを判定し、拒絶と判定された場合、ステップ2406で文字行を抽出し、ステップ2407で文字行を利用した帳票識別を行う。

## 【0094】

図24に示す処理は、前述のように、ある特徴を利用した帳票識別が拒絶された場合、別の特徴を利用した帳票識別を実行しているため、処理時間を短縮することができる。

## 【0095】

図26は処理センターと遠隔地にある支店や事務所とがネットワークに接続されて帳票識別を行うシステムの構成例を示すブロック図である。図26において、2600はネットワーク、2601はセンター側ステーション、2602は登録ステーション、2603、2614は作成ステーション、2610～2612は支店ステーション、2613は識別ステーションである。

## 【0096】

図26に示すシステムは、ネットワーク2600に、全国に数箇所あるセンター側ステーション2601と、遠隔地にある支店や事務所に設けられる支店支店ステーション2610～2612が接続されて構成されている。センター側ステーション2601には、帳票登録を行う登録ステーション2602、帳票を作成する作成ステーション2603が備えられている。また、支店ステーション2610～2612には、帳票画像の入力、識別を行う識別ステーション2613、支店での帳票作成を可能にする作成ステーション2614が備えられている。こ

のように構成されるシステムにおいて、ネットワーク2600を介して帳票識別用の辞書や印刷用の帳票ファイルが各支店ステーション2610～2612にセンター側ステーション2601から配布されるので、辞書などの保守を一括管理することができ、保守作業を軽減することができる。

## 【0097】

図21は本発明の他の実施形態による帳票画像識別システムの構成例を示すブロック図である。図21において、2110はセンター側ステーション、2111は遠隔地側ステーション、2112は帳票作成ステーション、2113は帳票ファイル、2114は帳票プリンタであり、他の符号は図1の場合と同一である。

## 【0098】

図21に示す本発明の他の実施形態による帳票画像識別システムは、通信ネットワーク100を介して辞書登録ステーション101、帳票識別ステーション106、帳票作成ステーション109、2112が接続されて構成されており、これらのステーションが連携して動作可能とされている。そして、辞書登録ステーション101、帳票作成ステーション109が、一点鎖線で囲ったセンター側ステーション2110として構成され、一方、帳票識別ステーション106、帳票作成ステーション2112が一点鎖線で囲った遠隔地にある各事務所に設置された遠隔地側ステーション2111として構成されている。

## 【0099】

辞書登録ステーション101は、帳票の種類を識別するために予め帳票画像の特徴を登録する。辞書登録ステーション101には、帳票画像を採取するスキャナ102、登録対象となった帳票画像を保管する登録対象帳票画像ファイル103、登録した帳票画像の特徴を保管する帳票辞書ファイル104が備えられている。また、辞書登録ステーション101は、表示装置105を有し、この表示装置105に、登録済み帳票画像やスキャナで採取した新たな帳票画像をその画面に表示し、対話的に登録作業を行う。

## 【0100】

帳票識別ステーション106は、スキャナ107から入力された帳票画像の種

類を、登録した帳票画像の特徴を保管する帳票辞書ファイル108を用いて識別する。帳票辞書ファイル108の内容は、辞書登録ステーション101側の帳票辞書ファイル104を複製したものであり、通信ネットワークを介して、常に最新のデータに更新することができる。なお、帳票識別に用いる帳票辞書ファイルとして、通信ネットワークを介して辞書登録ステーション側の帳票辞書ファイル104をアクセスして用いる構成でもよい。

#### 【0101】

帳票作成ステーション109は、帳票ファイル110と帳票プリンタ111とを備えており、帳票ファイル110に保管された帳票データに対して、帳票プリンタ111を用い帳票を紙に印刷する。帳票作成ステーション109は、帳票識別が容易になるよう帳票データを編集して、印刷し、辞書登録ステーション101において、スキャナ102で印刷した帳票の画像を採取し新規登録の試行を行うことができる。

#### 【0102】

帳票作成ステーション2112は、事務所側である遠隔地側ステーションに備えられており、帳票ファイル2113、帳票プリンタ2114が接続されている。そして、帳票ファイル2113に保管されている帳票は、事務所側で帳票プリンタ2114から印刷して配布する。このとき、帳票辞書ファイル108が更新され、最新の辞書ファイルが保管されていない場合、帳票辞書ファイル108と帳票ファイル2113との間の整合性がなくなり、前述で印刷して配布した帳票に対して帳票識別が拒絶される恐れがある。このため、帳票辞書ファイル108と帳票ファイル2113の整合性を検定する必要がある。

#### 【0103】

図25は帳票辞書ファイル108と帳票ファイル2113との間の整合性の検定を行う処理動作を説明するフローチャートである。

#### 【0104】

図25に示すフローにおいて、帳票作成ステーション2112は、まず、ステップ2500で印刷対象の帳票の種類を指定する。そして、ステップ2501で帳票辞書ファイル108を探索し、当該帳票辞書が保管されているか否かをステ

ップ2502で判定する。もし、当該帳票辞書が保管されている場合、ステップ2503で帳票ファイル2113から該当する帳票データを選択して、ステップ2504で帳票データを印刷する。一方、ステーション2502の判定で、もし、当該帳票辞書が保管されていなかった場合、ステップ2505で帳票辞書の送信要求をセンター側にある辞書登録ステーション101に送り、帳票辞書ファイル108を更新する。これにより、印刷した帳票に対応した帳票識別用辞書が保管されていることを保証することができる。

## 【0105】

## 【発明の効果】

以上説明したように本発明によれば、枠線の傾きや帳票縁の傾きより検出された帳票の傾き角を基に、枠の中心点座標に対して傾きの補正を行っているため、傾きを有する帳票に対しても高精度な帳票識別が可能である。また、本発明によれば、伸縮率を複数通りに仮定して入力点座標の拡大縮小を行っているため、特徴に欠落や偽特徴が発生している帳票が伸縮している場合にも、類似度を算出することができるという効果を得ることができる。

## 【0106】

さらに、本発明によれば、原点を一定間隔で横方向ならびに縦方向に平行移動して点同士の照合を行っているため、入力帳票に位置ずれがある場合にも、高精度に帳票識別を行うことができる。また、本発明によれば、枠線の種類が異なる帳票に対して、帳票毎あるいは帳票内の枠毎に枠線の種類の有効、無効を切り替えて帳票の種類を識別することができるため、帳票識別の信頼性の向上を図ることができ、かつ、記憶容量を低減することができるという効果を得ることができる。

## 【0107】

さらに、本発明によれば、入力特徴として枠の中心点を90度単位で回転させて照合を行っているため、スキャナに帳票が任意の向きに設定されていても帳票の識別が可能となり、また、罫線の分布特徴を利用する帳票識別と比較して信頼性の高い帳票の識別を行うことができる。

## 【図面の簡単な説明】

【図1】

本発明の一実施形態による帳票識別及び帳票登録を実行する帳票識別システムの構成例を示すブロック図である。

【図2】

辞書登録ステーションでの帳票識別のための辞書登録の処理動作を説明するフローチャートである。

【図3】

図2により説明した辞書登録における対話操作を行うための画面表示例を説明する図である。

【図4】

帳票識別ステーションでの入力帳票の特徴抽出処理と帳票識別処理との動作を説明するフローチャートである。

【図5】

図4のステップ401での帳票識別処理の動作を説明するフローチャートである。

【図6】

図5のステップ501での探索照合処理の動作を説明するフローチャートである。

【図7】

図5のステップ604での点照合処理の動作を説明するフローチャートである。

【図8】

図7のステップ702での類似度抽出処理の動作を説明するフローチャートである。

【図9】

図8のステップ802での投票数抽出処理の動作を説明するフローチャートである。

【図10】

図4のステップ400での枠抽出処理の動作を説明するフローチャートである。

【図11】

帳票識別ステーション106での入力帳票の特徴抽出処理と帳票識別処理と他の動作例を説明するフローチャートである。

【図12】

図4のステップ400での枠抽出処理について説明する帳票識別の特徴としての枠の中心点について説明する図である。

【図13】

帳票辞書ファイルに格納されている辞書点について説明する図である。

【図14】

図9のステップ903での入力点と辞書点との最小距離を算出する方法を説明する図である。

【図15】

図6のステップ602、603で示した座標値をX、Y方向へずらした点照合について説明する図である。

【図16】

枠の特徴として用いる枠線の種類について説明する図である。

【図17】

図8のステップ802での投票数抽出の処理において、点の属性を利用する場合の処理動作を説明するフローチャートである。

【図18】

図5のステップ500での枠座標の90度回転の処理について説明する図である。

【図19】

図8のステップ802での投票数抽出の処理において、点の属性を有効または無効の何れかに切り替えて投票数を求める場合の処理動作を説明するフローチャートである。

【図20】

帳票識別に用いる帳票画像の特徴例について説明する図である。



【図 2 1】

本発明の他の実施形態による帳票画像識別システムの構成例を示すブロック図である。

【図 2 2】

帳票画像の特徴として、枠、罫線、文字行矩形を用いて帳票の識別を行う処理動作の例を説明するフローチャートである。

【図 2 3】

図 2 2 のステップ 2 2 0 6 での帳票識別結果の複合判定の処理動作を説明するフローチャートである。

【図 2 4】

3 種の帳票の特徴を用いた帳票識別の処理動作の他の例を説明するフローチャートである。

【図 2 5】

帳票辞書ファイル 1 0 8 と帳票ファイル 2 1 1 3 との間の整合性の検定を行う処理動作を説明するフローチャートである。

【図 2 6】

処理センターと遠隔地にある支店や事務所とがネットワークに接続されて帳票識別を行うシステムの構成例を示すブロック図である。

【符号の説明】

- 1 0 0、2 6 0 0 ネットワーク
- 1 0 1 辞書登録ステーション
- 1 0 2、1 0 7 スキャナ
- 1 0 3 登録対象帳票画像ファイル
- 1 0 4、1 0 8 帳票辞書ファイル
- 1 0 5 表示装置
- 1 0 6 帳票識別ステーション
- 1 0 9、2 1 1 2 帳票作成ステーション
- 1 1 0、2 1 1 3 帳票ファイル
- 1 1 1、2 1 1 4 帳票プリンタ

2110、2601 センター側ステーション

2111 遠隔地側ステーション

2602 登録ステーション

2603、2614 作成ステーション

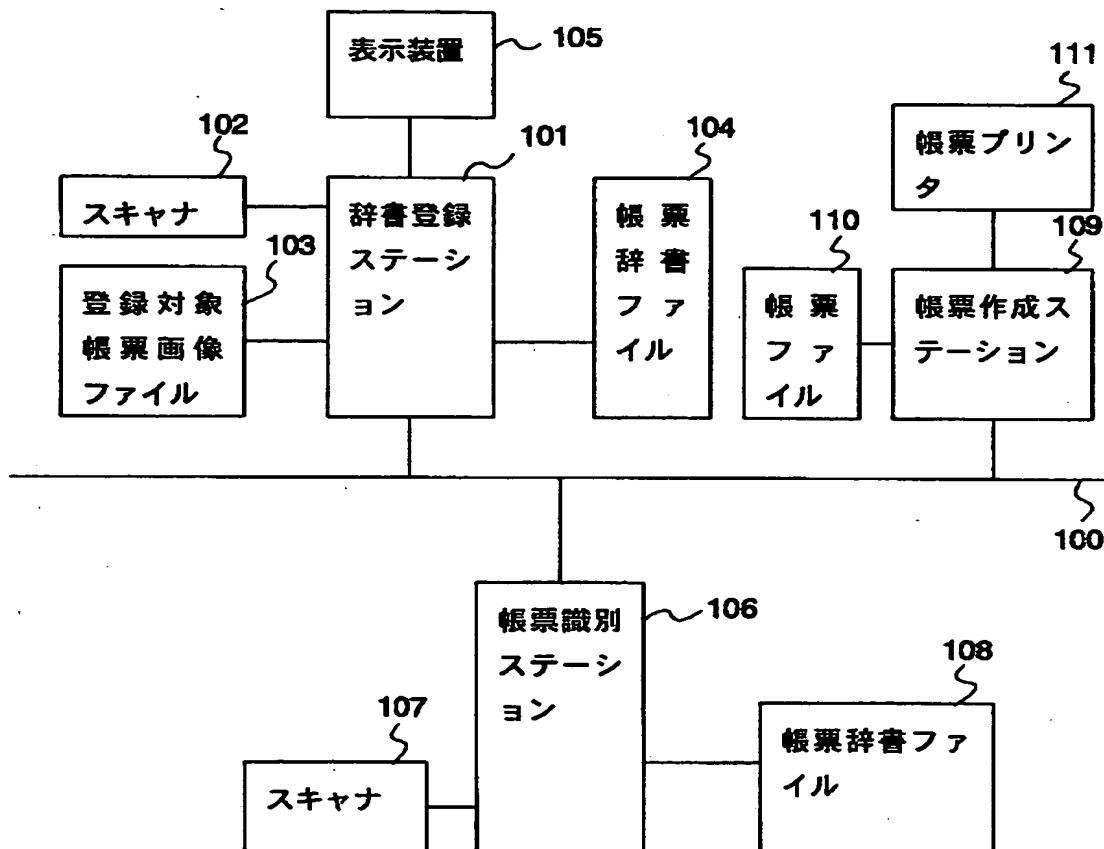
2610～2612 支店ステーション

2613 識別ステーション

【書類名】 図面

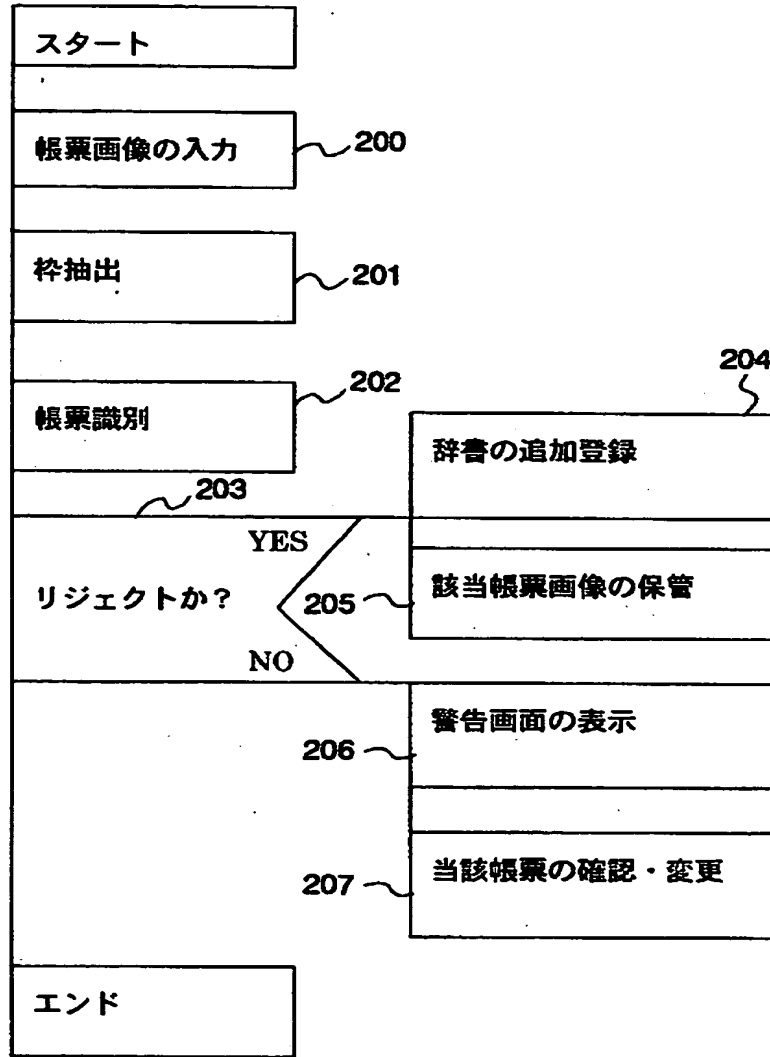
【図1】

図1



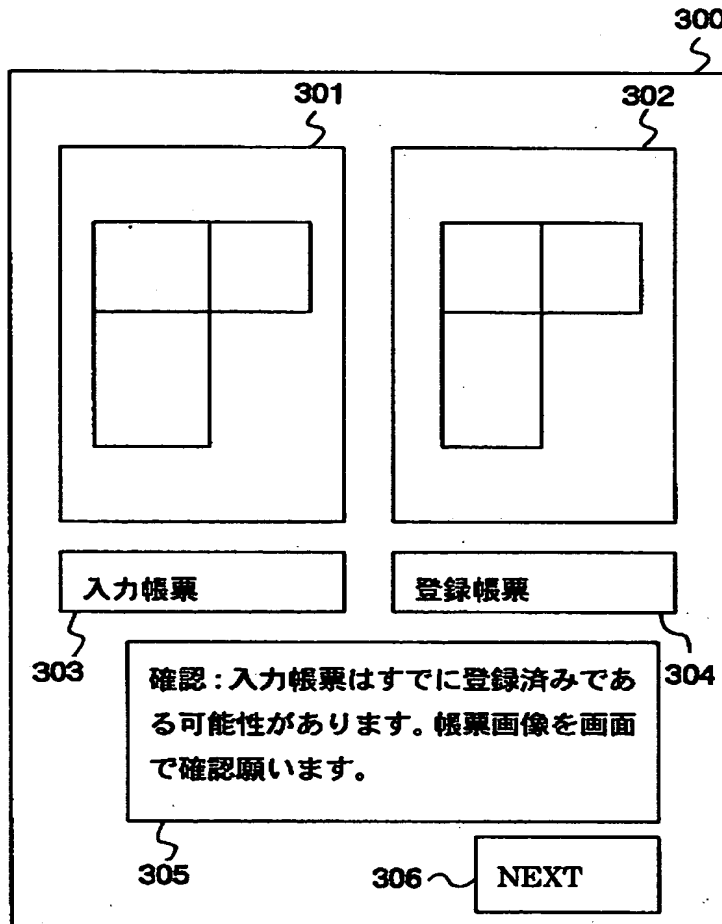
【図 2】

図 2



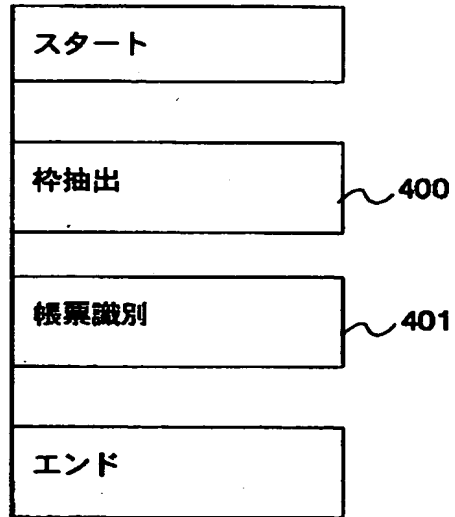
【図3】

図3



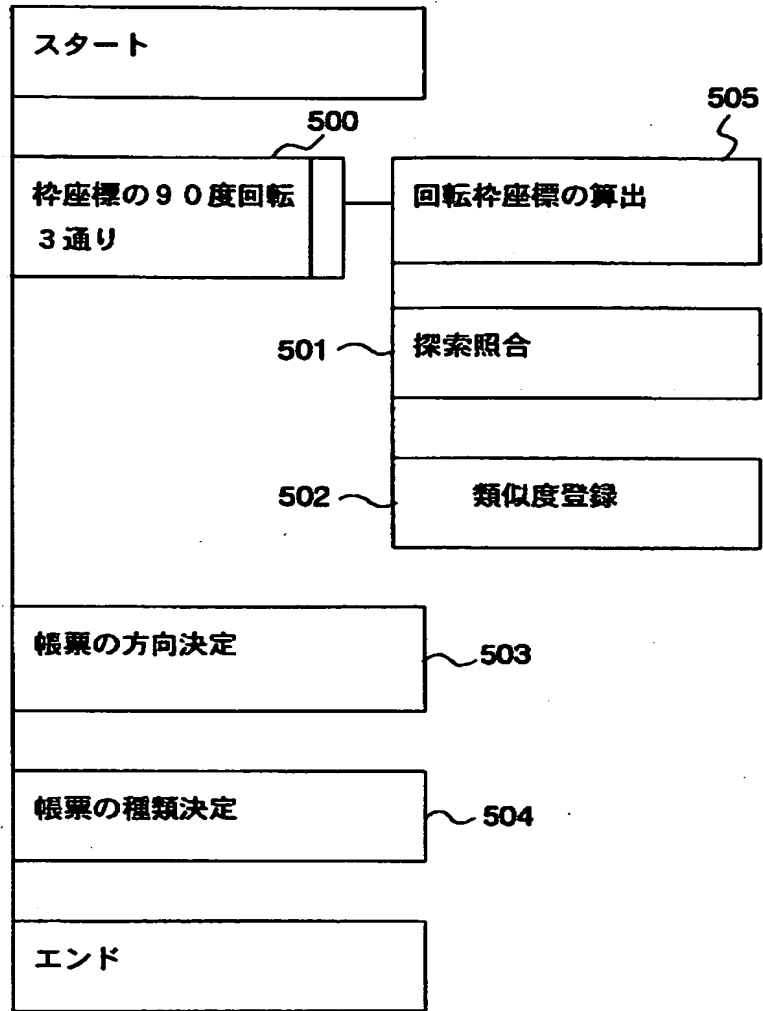
【図 4】

図 4



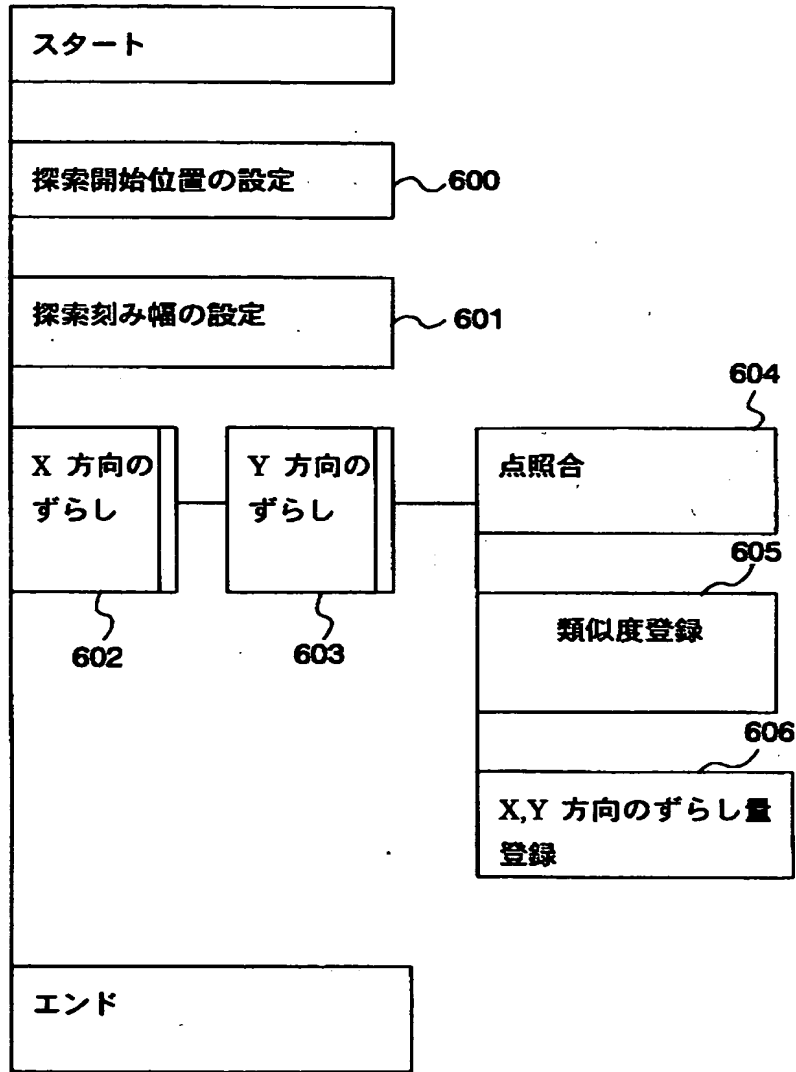
【図5】

図5



【図 6】

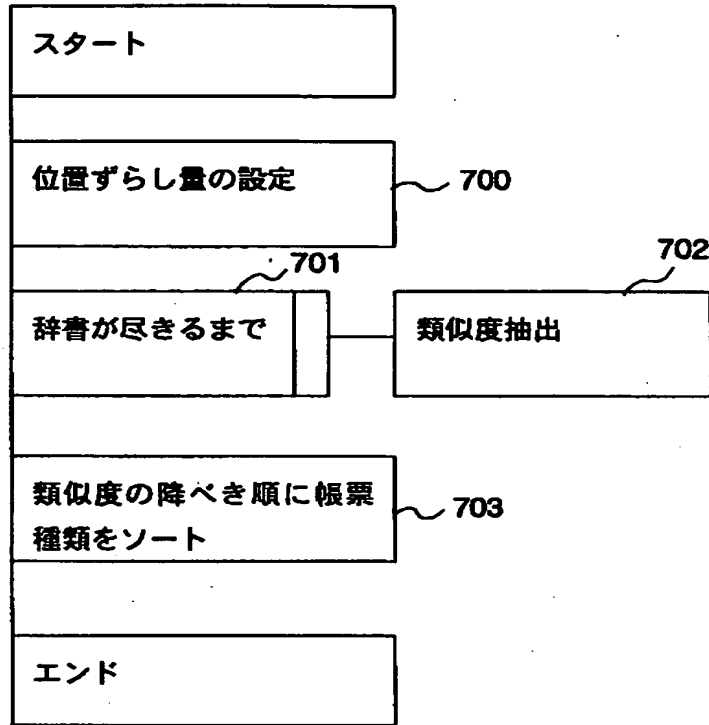
図 6





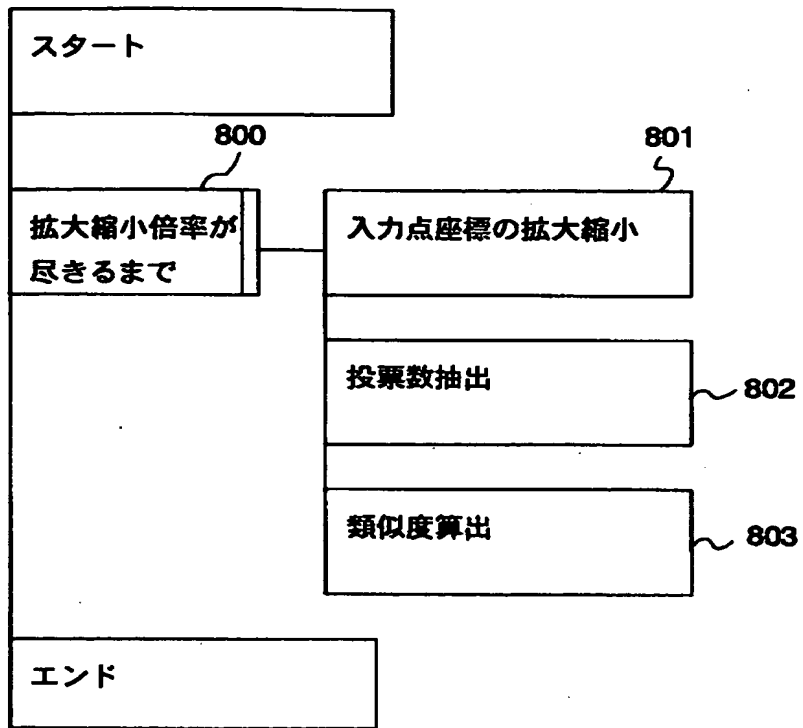
【図7】

図7

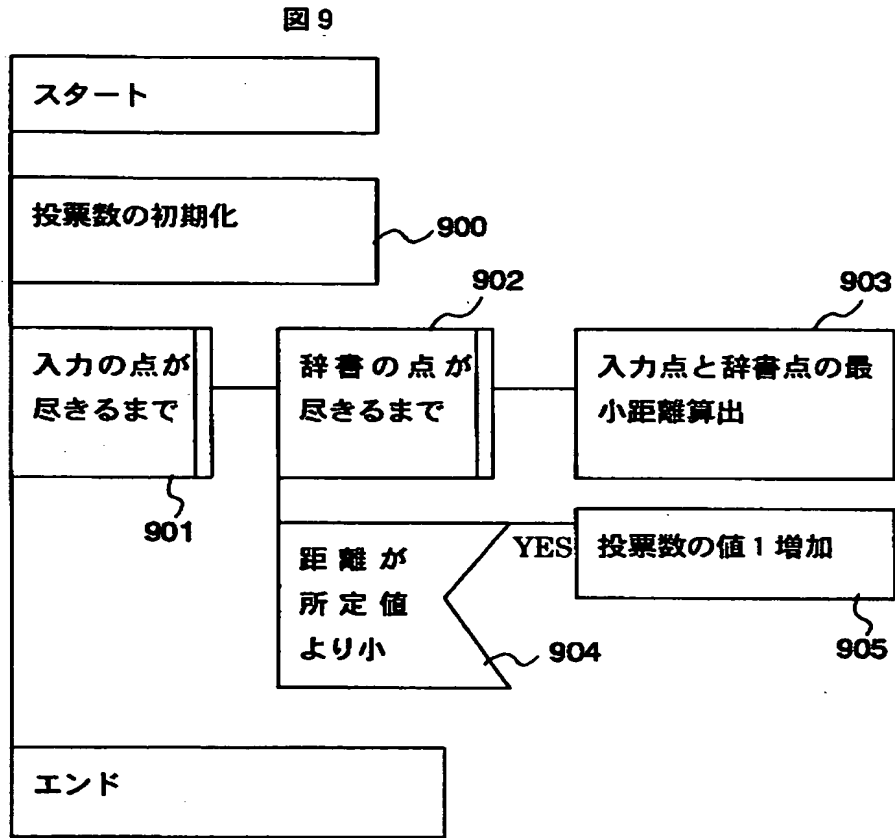


【図8】

図8

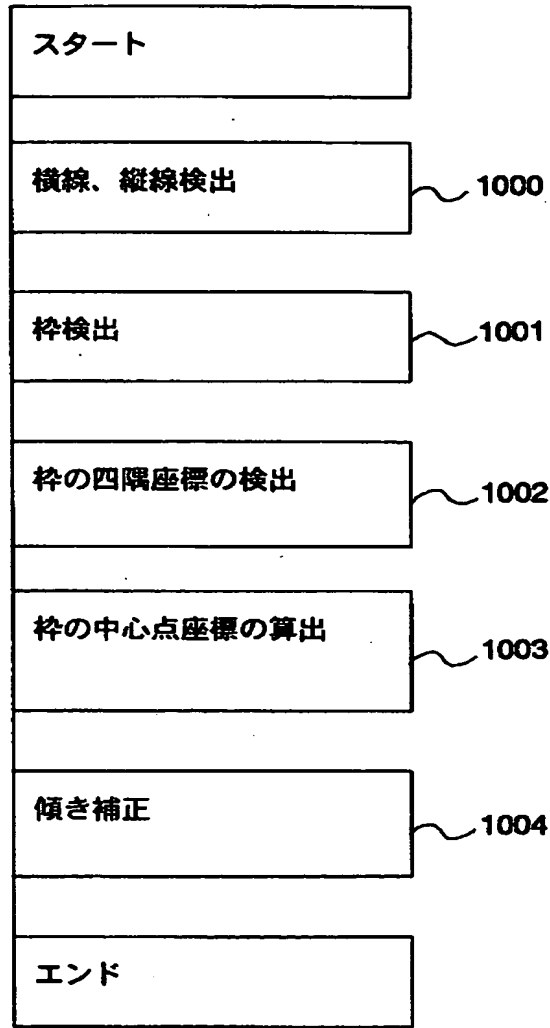


【図9】



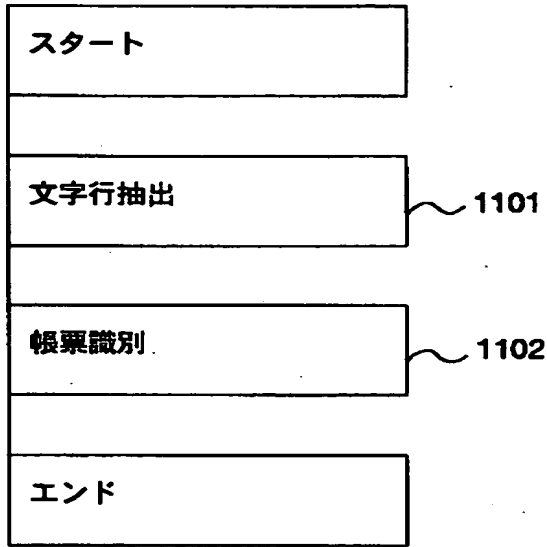
【図10】

図10



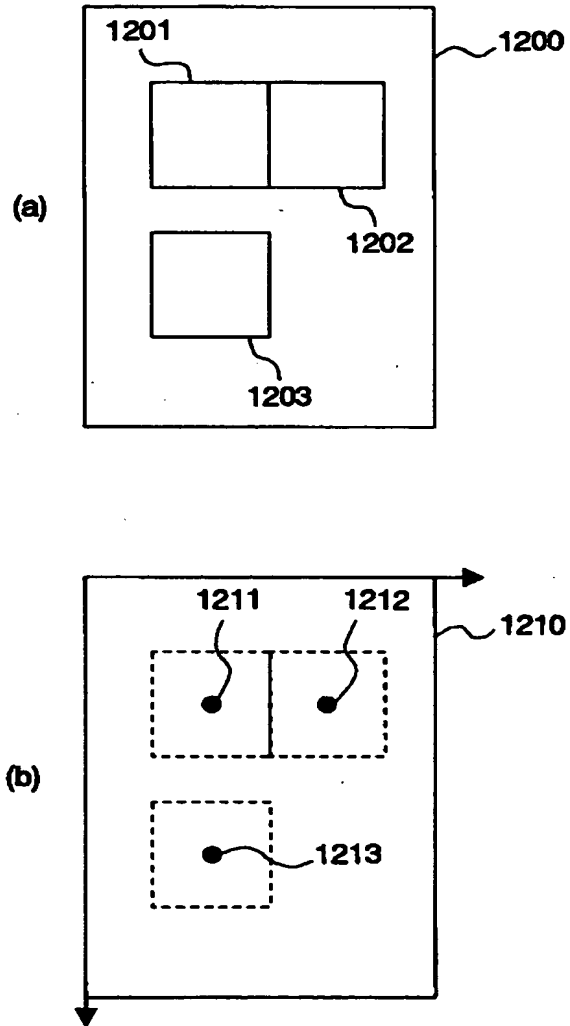
【図 11】

図 11



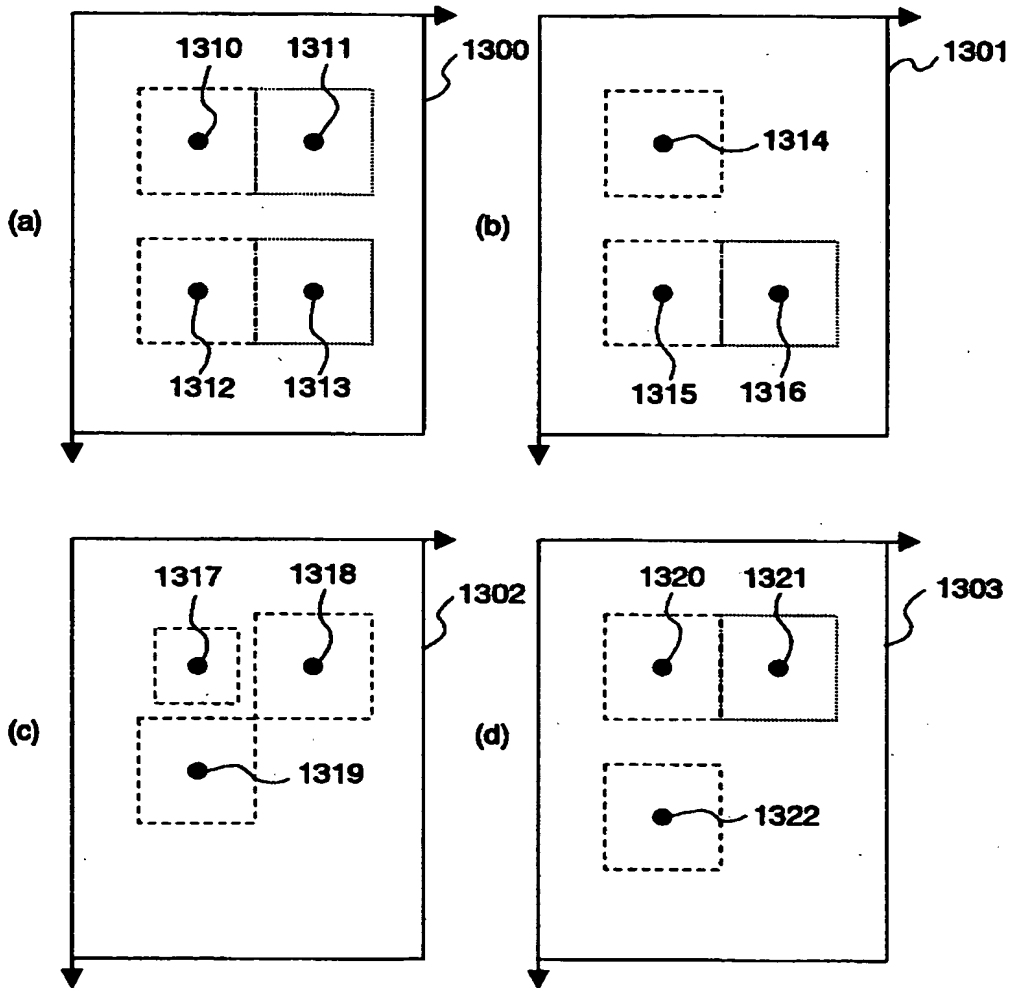
【 図 1 2 】

图 1 2

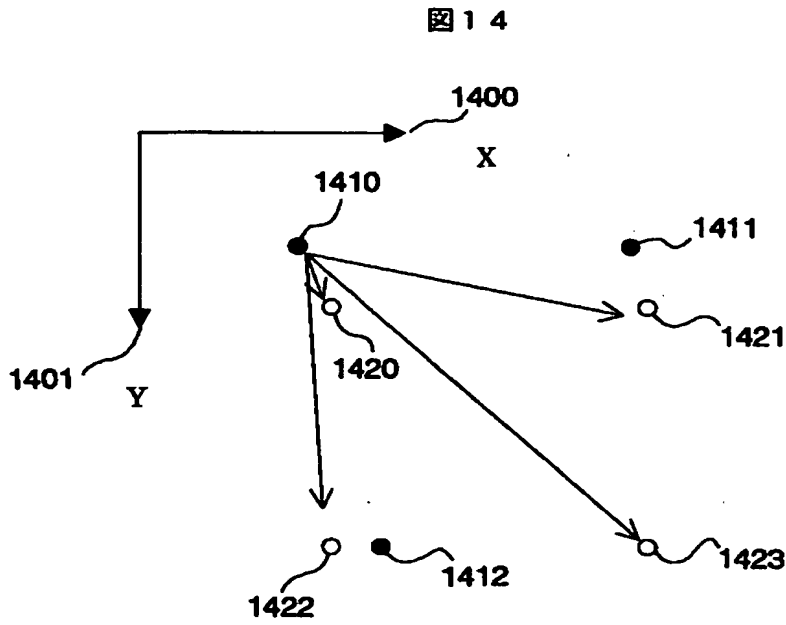


【図 13】

図 13

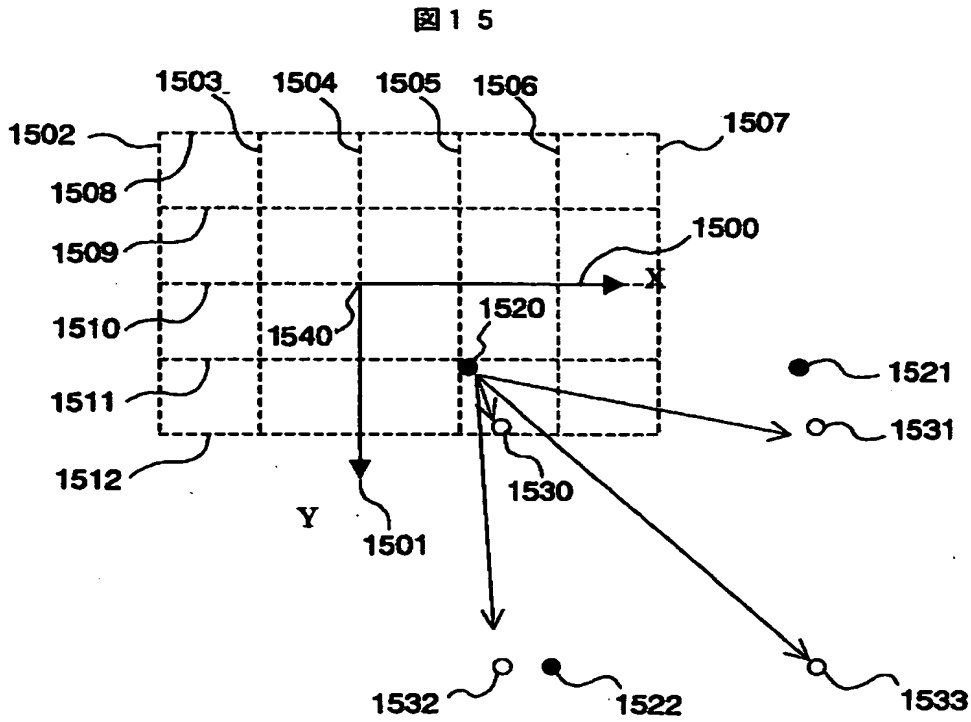


【图 14】



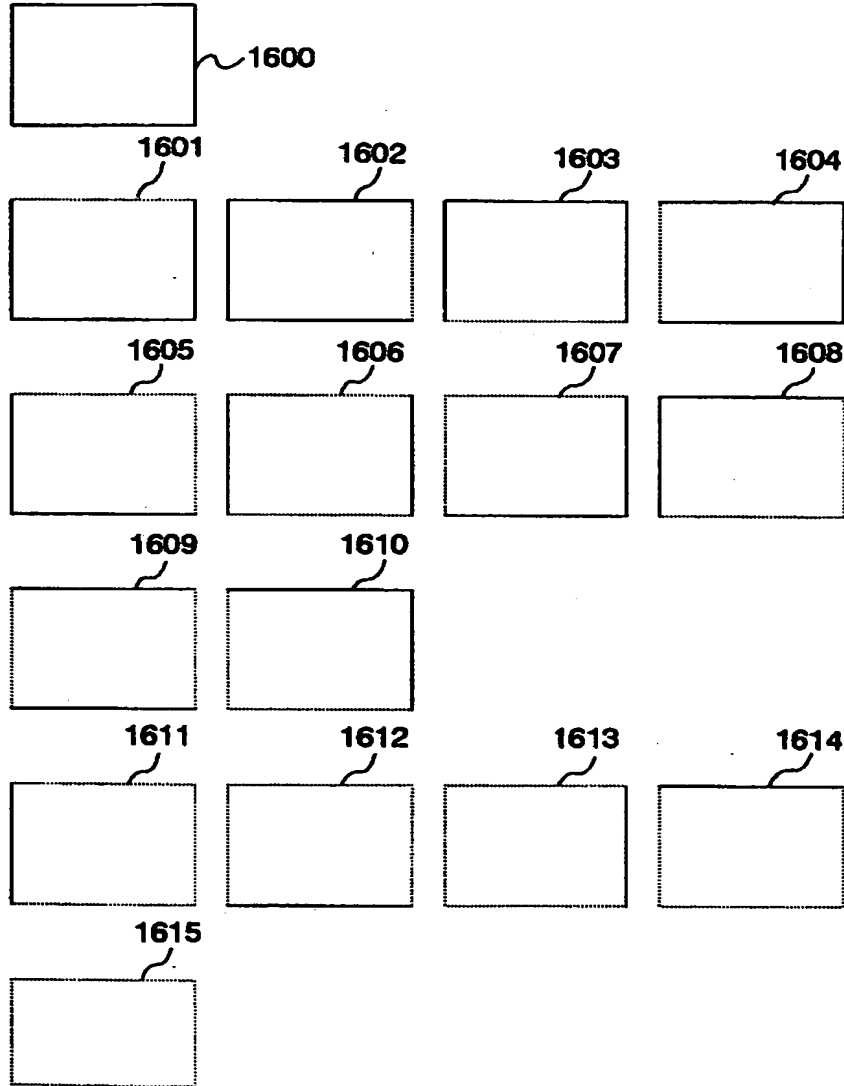


【図15】



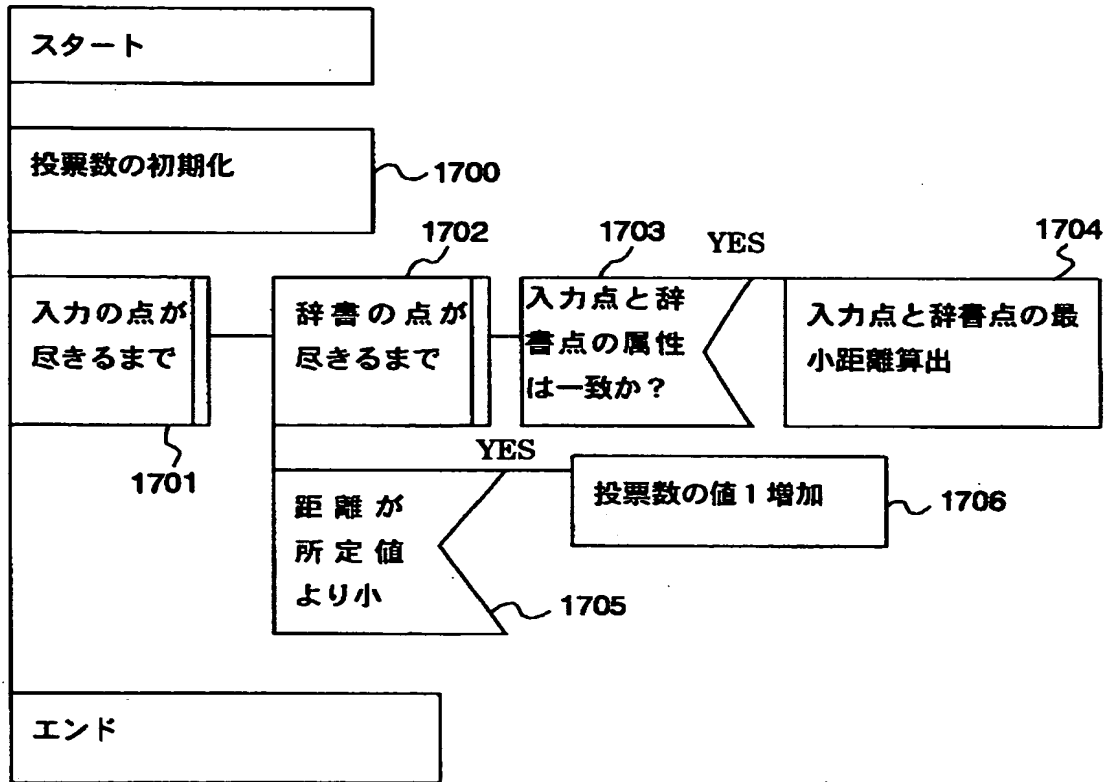
【図 16】

図 16



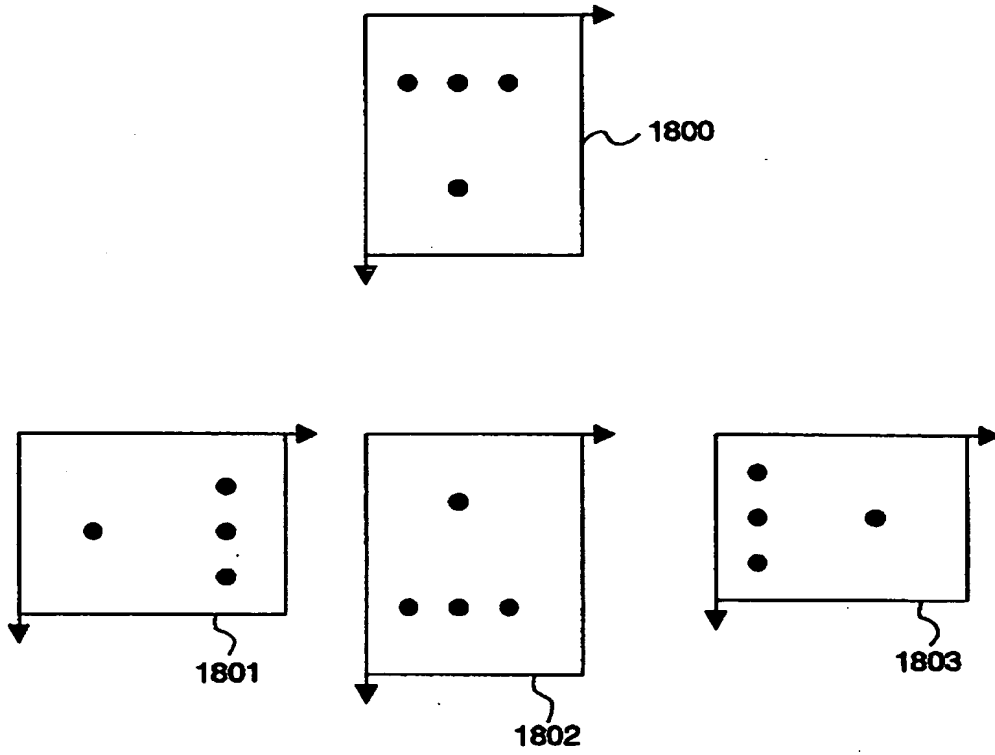
【図 17】

図 17



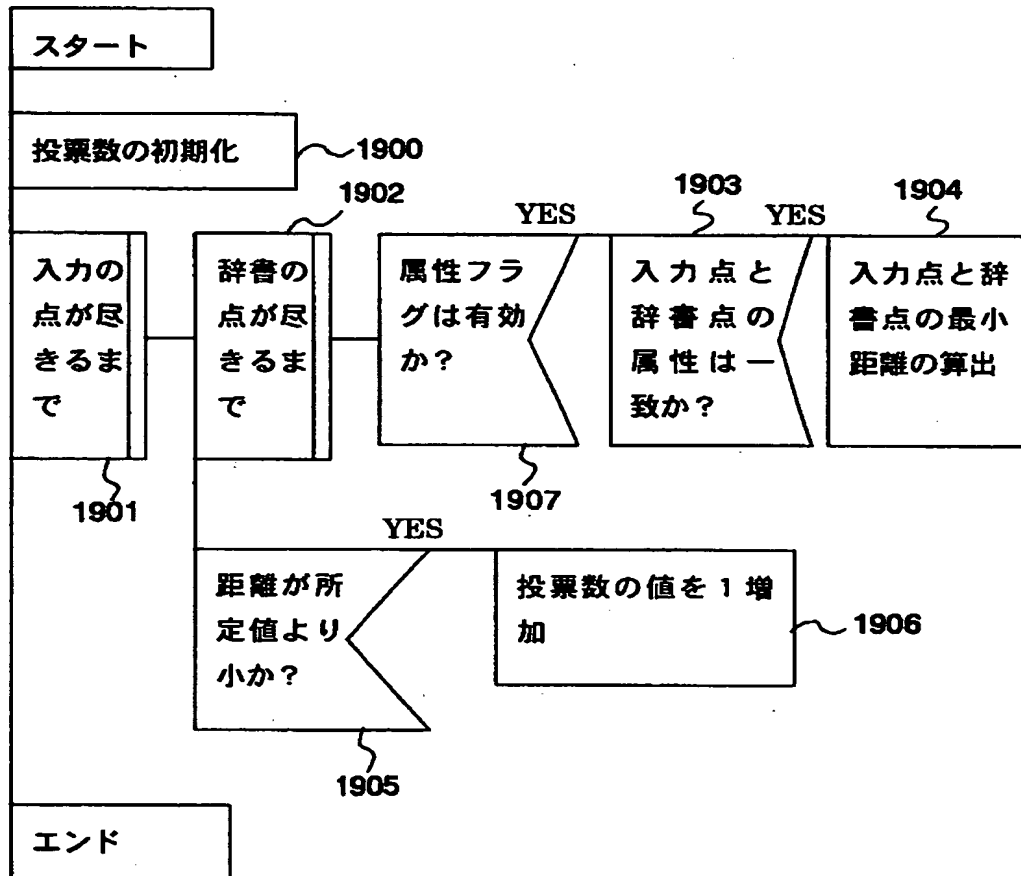
【図 18】

図 18



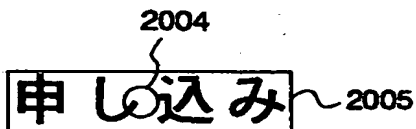
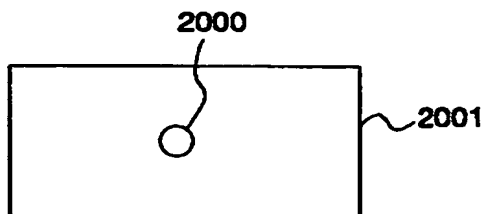
【図19】

図19



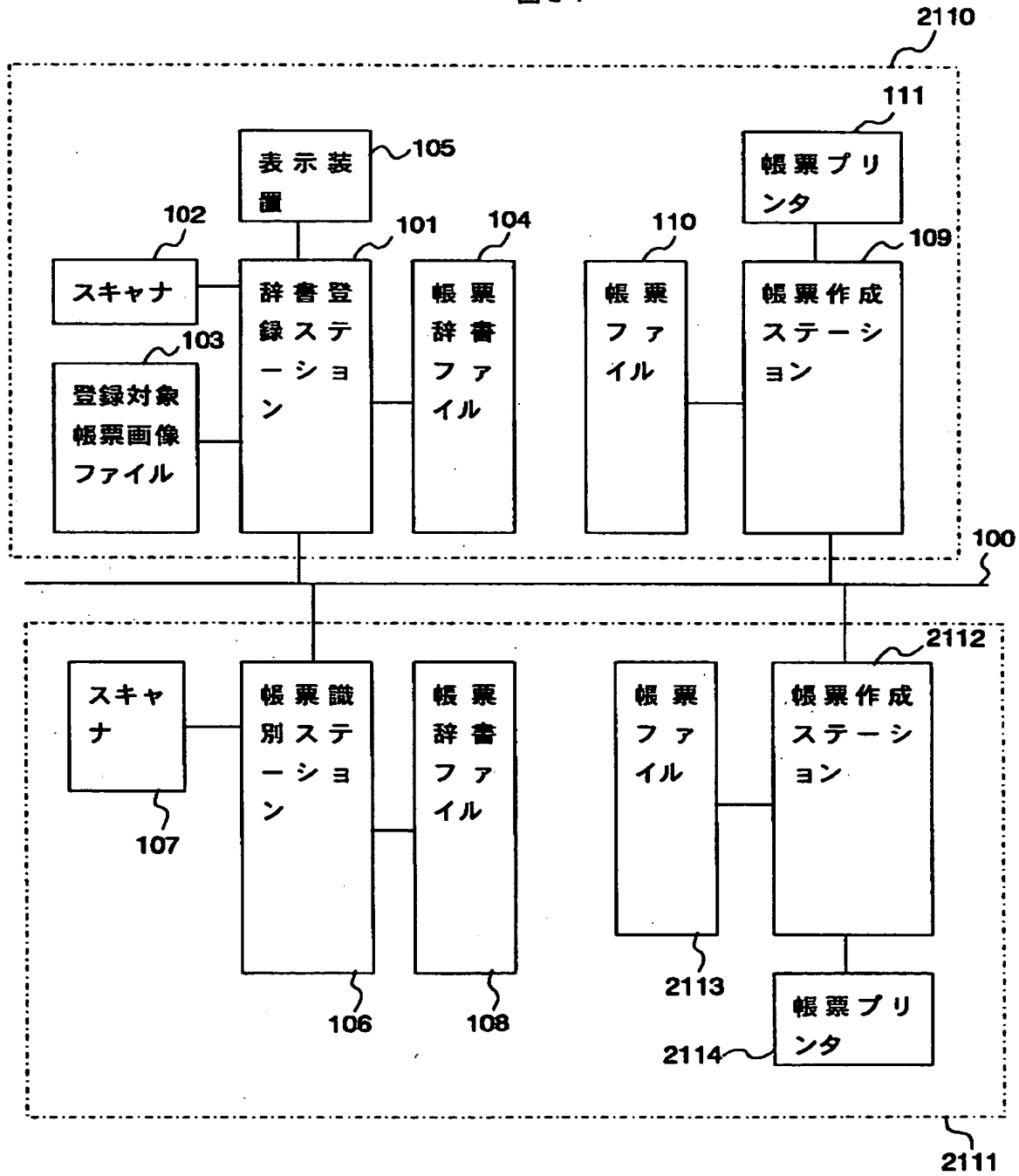
【図 2 0】

図 2 0



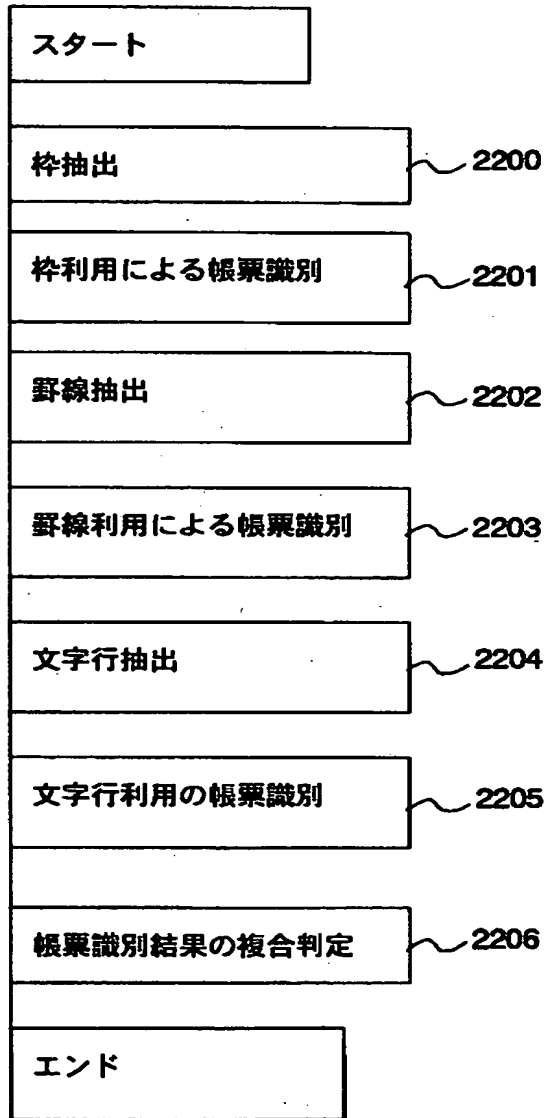
【図21】

図21



【図 2 2】

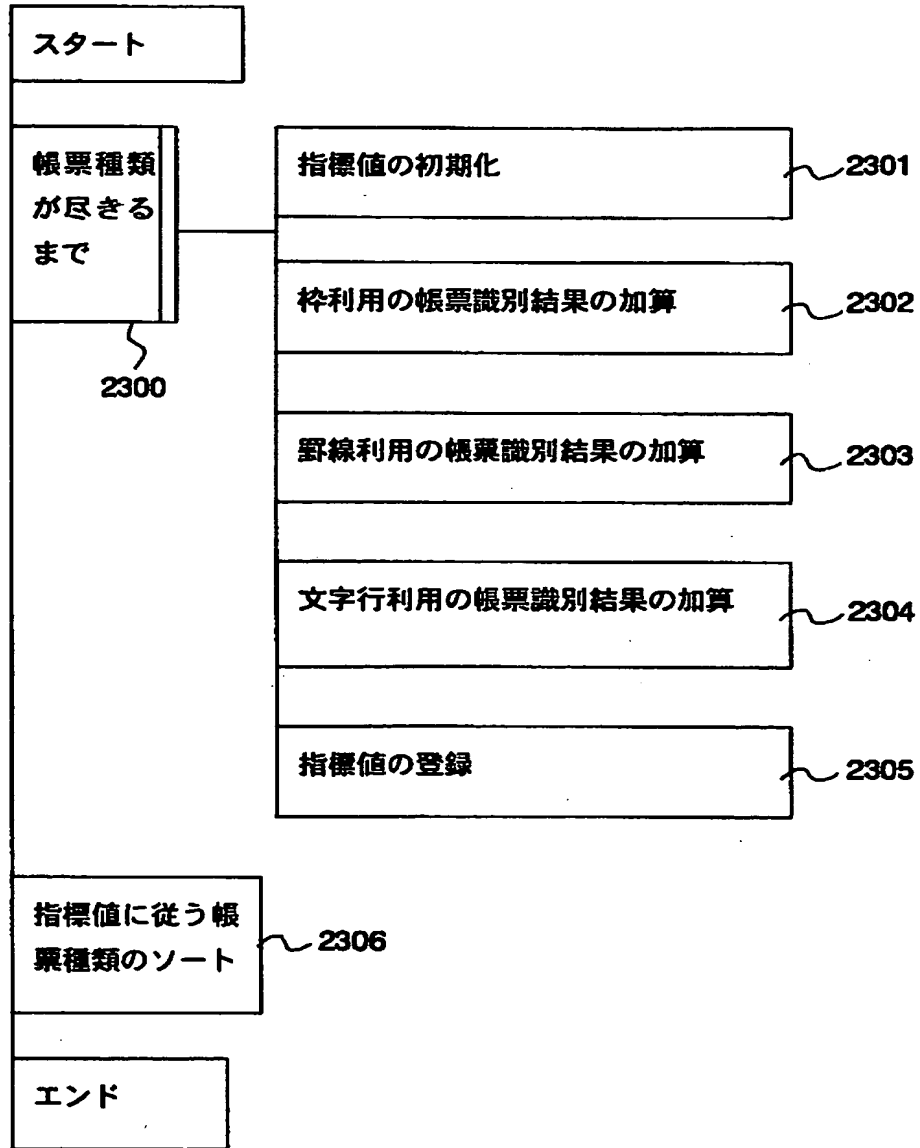
図 2 2





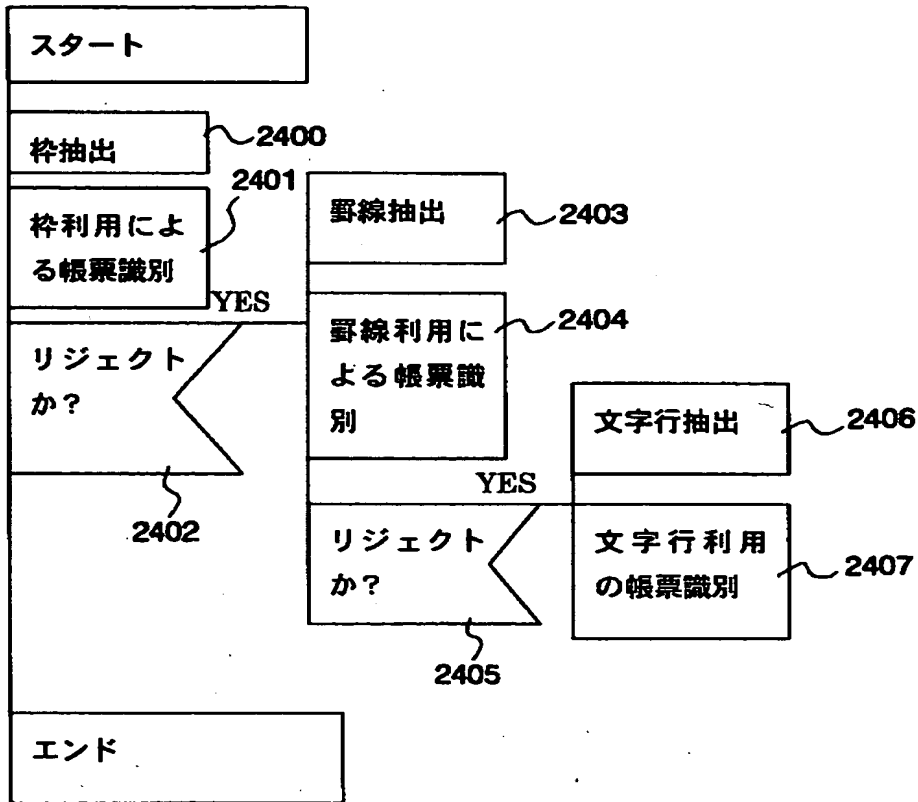
【図 23】

図 23



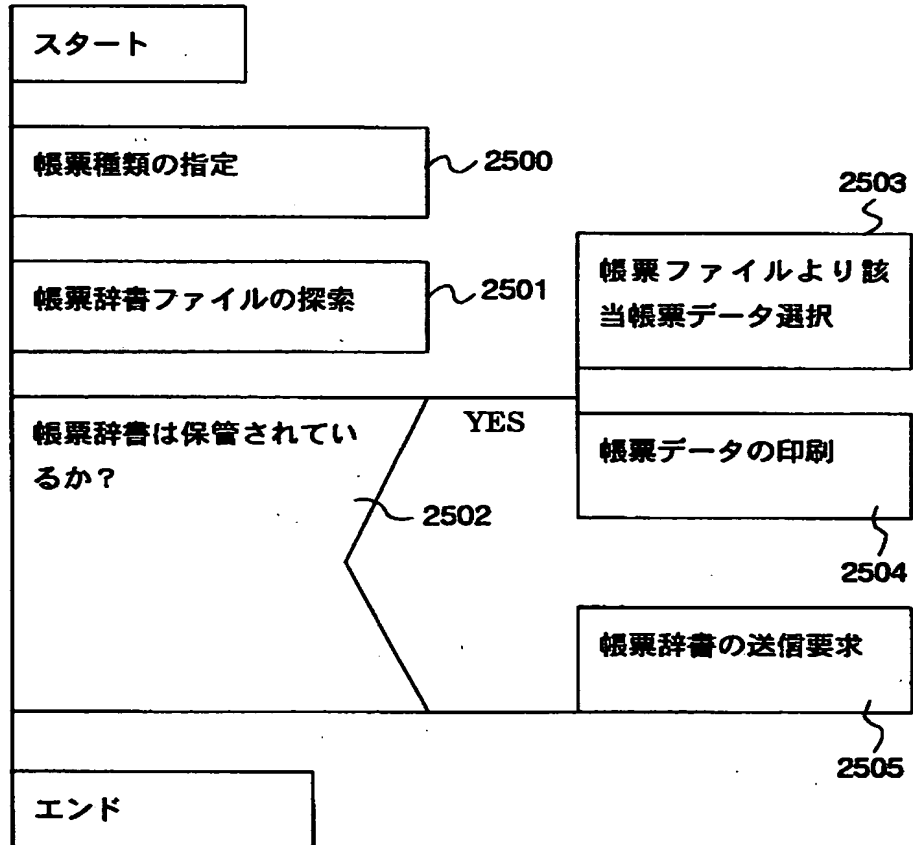
【図 24】

図 24



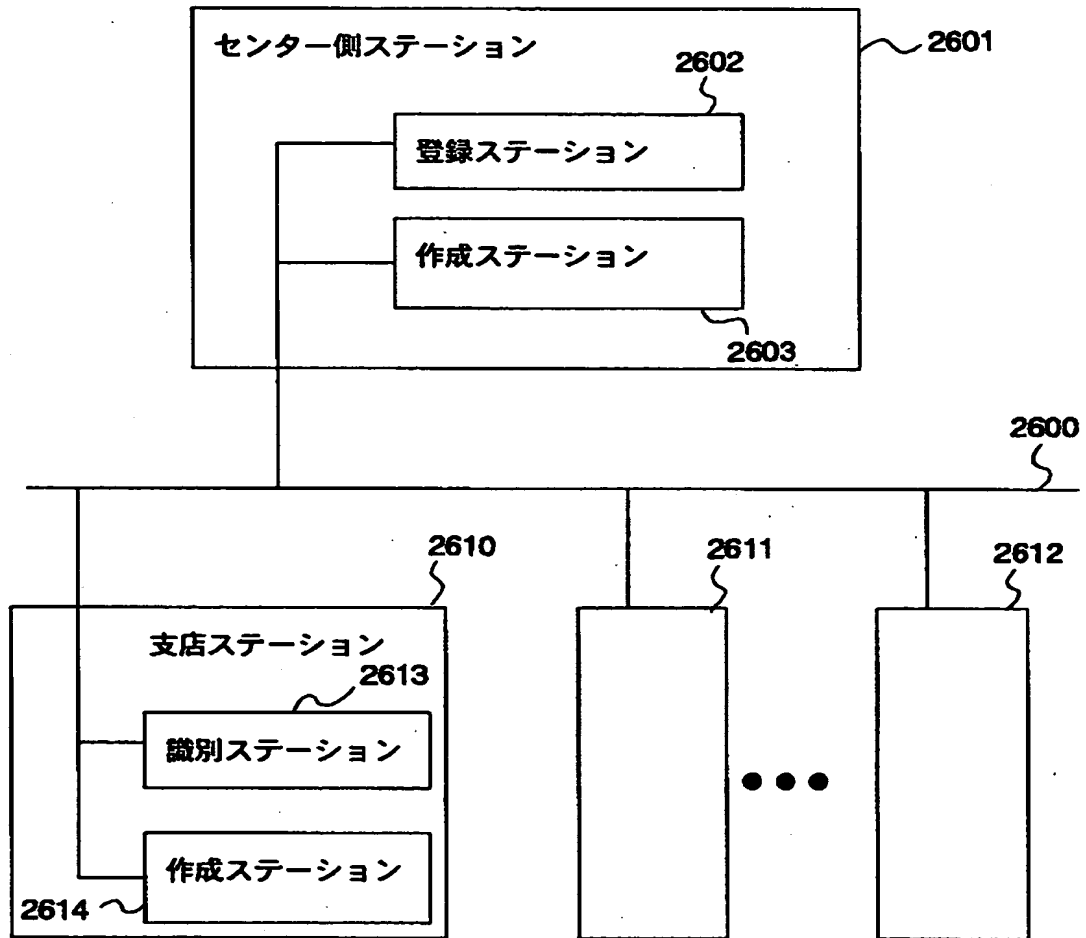
【図25】

図25



【図 26】

図 26



【書類名】 要約書

【要約】

【課題】 スキャナから入力される帳票の向きが異なっても、また、帳票が伸縮を有するものであっても、誤りなく帳票識別可能とする。また、予め帳票の種類を登録する際に、帳票の変更修正を行う構成とを可能とする。

【解決手段】 帳票から特徴となる入力点を抽出し、伸縮、回転、傾きの補正を入力点に対して行い、予め登録している辞書点との距離の最小値に基づいて帳票の種類を識別する。また、帳票種類の登録時、類似帳票に対して警告画面表示することを可能にする。システムは、通信ネットワーク100を介して辞書登録ステーション101、帳票識別ステーション106、帳票作成ステーション109が接続されて構成される。各ステーションは、お互いに連携して動作し、入力帳票の識別、辞書への登録を行う。

【選択図】 図1

出願人履歴情報

識別番号 [000005108]

1. 変更年月日 1990年 8月31日  
[変更理由] 新規登録  
住所 東京都千代田区神田駿河台4丁目6番地  
氏名 株式会社日立製作所

出願人履歴情報

識別番号 [000233033]

1. 変更年月日 1990年 8月24日  
[変更理由] 新規登録  
住所 神奈川県小田原市国府津2880番地  
氏名 日立コンピュータ機器株式会社
  
2. 変更年月日 2001年 7月24日  
[変更理由] 住所変更  
住所 神奈川県足柄上郡中井町境781番地  
氏名 日立コンピュータ機器株式会社